Understanding Google: Search Engines and the Changing Nature of Access, Thought and Knowledge within a Global Context

Submitted by Richard Graham to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in English
In December 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……

Abstract

This thesis explores the impact of Google's search engine within contemporary digital culture. Search engines have been studied in various disciplines, for example information retrieval, computer science, law, and new media, yet much of this work remains fixed within disciplinary boundaries. The approach of this thesis is to draw on work from a number of areas in order to link a technical understanding of how search engines function with a wider cultural and philosophical context. In particular, this thesis draws on critical theory in order to attend to the convergence of language, programming, and culture on a global scale.

The chapter outline is as follows. Chapter one compares search engine queries to traditional questions. The chapter draws from information retrieval research to provide a technical framework that is brought into contact with philosophy and critical theory, including Plato and Hans-Georg Gadamer. Chapter two investigates search engines as memory aids, deploying a history of memory and exploring practices within oral cultures and mnemonic techniques such as the *Ars Memoria*. This places search engines within a longer historical context, while drawing on contemporary insights from the philosophy and science of cognition. Chapter three addresses Google's Autocomplete functionality and chapter four explores the contextual nature of results in order to highlight how different characteristics of users are used to personalise access to the web. These chapters address Google's role within a global context and the implications for identity and community online. Finally, chapter five explores how Google's method of generating revenue, through advertising, has a social impact on the web as a whole, particularly when considered through the lens of contemporary Post-Fordist accounts of capitalism. Throughout, this thesis develops a framework for attending to algorithmic cultures and outlines the specific influence that Google has had on the web and continues to have at a global scale.

Table of Contents

Acknowledgements

The research of this thesis was supported through funding by the Arts and Humanities Research Council's Digital Humanities Doctoral Award (Reference 1099). I am grateful for the financial support that enabled my doctoral study.

I am sincerely indebted to my supervision team, Regenia Gagnier and Laura Salisbury, for their expertise and encouragement throughout the thesis process.

My deepest thanks go to the University of Exeter's English faculty, not only for the many supportive conversations during the thesis but also for the inspirational teaching during my undergraduate and master's degrees. I particularly owe my decision to follow academia to the rousing lectures and seminars of Matt Hayler and Paul Williams. I also wish to thank my students, particularly the first-year critical theory students of Approaches to Criticism. Their enthusiasm and often vocal bemusement is a continual reminder of the joys of academic study.

I would like to thank all of those whom I sailed with on Semester at Sea for the many enlightening conversations. I am indebted to Victor Luftig for the opportunity and am grateful for his wisdom and sincerity. My thanks also go to Alex Nalbach and Mark Brill for their infectious enthusiasm for teaching and research.

I have been exceptionally lucky for the friendships that have helped me through this period and in particular the patience, kindness and thoughtfulness of James Tyndall, Jack Griffiths and Ryan Sweet.

Finally, this thesis is dedicated to the following people:

Anne and **Keith Graham,** without your longstanding support and resilience I would not be where I am now. You have enabled this journey in countless ways.

Matt Hayler, I could not wish for a better mentor and friend. The origins of this research and interest in all things digital stem from our many conversations.

Regenia Gagnier, for as long as I write, your approval will always be my watermark of quality. Your directness and encouragement have been invaluable and I will forever treasure the many pieces of advice you have given me over the years.

Zoe Bulaitis, you have been a significant part of every step of this journey. None of this would have been possible without your support, patience and enthusiasm.

Introduction

This thesis explores the impact of search engines within contemporary digital culture and, in particular, focuses on the social, cultural, and philosophical influence of Google. Search engines are deeply enmeshed with other recent developments in digital culture; therefore, in addressing their impact these intersections must be recognised, while highlighting the technological and social specificity of search engines. Also important is acknowledging the way that certain institutions, in particular Google, have shaped the web and wider culture around a particular set of economic incentives that have far-reaching consequences for contemporary digital culture. This thesis argues that to understand search engines requires a recognition of its contemporary context, while also acknowledging that Google's quest to "organize the world's information and make it universally accessible and useful" is part of a much older and broader discourse. Balancing these two viewpoints is important; Google is shaping public discourse on a global scale with unprecedentedly extensive consequences. However, many of the issues addressed by this thesis would remain centrally important even if Google declared bankruptcy or if search engines were abandoned for a different technology. Search engines are a specific technological response to a particular cultural environment; however, their social function and technical operation are embedded within a historical relationship to enquiry and inscription that stretches back to antiquity.

This thesis addresses the following broad research questions, while at each stage specifically addressing the role and influence of search engines: how do individuals interrogate and navigate the world around them? How do technologies and social institutions facilitate how we think and remember? How culturally situated

is knowledge; are there epistemological truths that transcend social environments? How does technological expansion fit within wider questions of globalisation? How do technological discourses shape the global flows of information and capital? These five questions map directly onto the five chapters of this thesis.

Much of the existing study of search engines has been focused on small-scale evaluation, which either addresses Google's day-by-day algorithmic changes or poses relatively isolated disciplinary questions. Therefore, not only is the number of academics, technicians, and journalists attending to search engines relatively small, given the centrality of search engines to digital culture, but much of the knowledge that is produced becomes outdated with algorithmic changes or the shifting strategies of companies. This thesis ties these focused concerns to wider issues, with a view to encourage and facilitate further enquiry.

1.0 Google's Dominance

While this thesis does address search engines other than Google, directly in chapter one and indirectly throughout, a representative portrayal of the history of the search engine market must place Google at the centre of that picture, for a number of reasons. Google established the main standards central to today's conception of what search engines represent and how they function. This has shaped the overall direction of search engines significantly; there have been many approaches to providing access to webpages, but Google's model has set the standard, which now all other engines follow. Globally, across all platforms, Google have a search engine market share of 92.01% as of July 2017 (see "Search engine market share Worldwide, July 2017"). To use an ecological metaphor, the narrative is a story of many web search engines, around 70 or more, all growing up towards the light and eventually being plunged into darkness by Google's overwhelming canopy. The nature of Google's monopoly is not an ancillary part of this history; this is because the greater number of individuals who use a particular search engine, the more informational and financial resources it has to shape its results. In addition, the hegemony of a particular company allows them to fix the cultural expectations for search engine results, which reinforces their position. Therefore, Google's success has provided them significant resources which they have invested in ways that consolidate their dominance. For example, Google's entry into the mobile industry, Android, as well as deals with other companies, enabled Google's search engine to become the default on the vast majority of mobile devices. This means that as the main way to access the web shifts from desktop to mobile, Google's dominance grows even further. For example, Google's desktop market share in the United

States is 77.45%, whereas its mobile market share in the US is 96.2% (see “Mobile Search Engine Market Share in the United States of America, July 2017”). In 2015, Google announced that “more Google searches take place on mobile devices than on computers in 10 countries including the US and Japan” (Dischler) and in 2016, Search Engine Land’s Greg Sterling estimated that in the US overall mobile search, which includes both smartphones and tablets, had risen to “58 percent of overall search query volume” (“Report: Nearly 60 Percent”). During the last twenty years, Google’s voice has drowned out the sound of its competitors. If the market remains on its current trajectory, the search engine narrative of the next twenty years will be in the form of a soliloquy.

Simply establishing an infrastructure and building up an index of the web, a prerequisite for any kind of search engine, is hugely costly. The only index that covers a large enough portion of the web to be comparable to Google’s is maintained by Microsoft. Unsuccessful public, as well as private, attempts have been made; in 2005, a joint German-French project (named Quaero) was established in order to build a European search engine. The venture failed before even managing to develop an alternative engine (see Winkler and European Commission), let alone maintain one that would provide enough of an incentive for users to switch from Google. The research and development costs, provided by the European Commission, were €199 million for the French and €120 million for the German development teams. Without an independent index of the web, any attempts to provide alternative search engine rankings require building and maintaining an index from scratch. With every passing day, the associated costs of building up an index of the web rise as the web grows larger and the opportunity to challenge the status quo shrinks as Google’s profits increase. That in the search engine market, Microsoft are

the underdog, is a telling sign. Although establishing an online landscape of transparency and accountability is overwhelmingly in the global public interest, after Quaero, another attempt at producing a publicly funded search engine is unlikely.

In almost every country, Google have a monopolistic share of the search engine market. There are some exceptions, such as Baidu in China, but these examples usually represent their own monopolies; the only exception is the fairly equal market share, between Yandex and Google, in Russia (although, for mobile share Google have pulled ahead to 65.51% (see “Mobile Search Engine Market Share in the Russian Federation, July 2017”)). However, in many countries, Google's share of the market sits between 90% and 100%. Because alternatives are the exception, rather than the rule, this thesis does not compare the results of different search engines. This could be a topic for further study, although if trends continue as they are, the real-world relevance diminishes daily. Another important point to note is that there are fewer alternatives to Google than might be assumed, as Dirk Lewandowski argues,

> many providers of what may appear to be a search engine are simply services that access the data of another search engine, representing nothing more than an alternative user interface to one of the more well-known engines, and in many cases, that turns out to be Google. (“Why We Need an Independent Index of the Web” 51)

For example, Yahoo! is powered by Microsoft’s Bing search engine and various meta-search engines DuckDuckGo and DogPile rely heavily on Google’s results. Even Bing does not necessarily represent an alternative to Google’s results; in 2011,

conclusive reports both conducted by Google and outside institutions found evidence that Bing was directly copying Google's results wholesale, claiming them as their own, as will be discussed in more depth in chapter four. Whether or not the search engine market naturally tends toward a monopoly, as is discussed by Siva Vaidhyanathan in *The Googlization of Everything (And Why We Should Worry)*, is not discussed in this thesis, although the nature of that debate has wide ramifications for many areas of digital culture. There are specific technical innovations that led to Google's early rise to power and these are discussed in chapter one, as well as later modifications, such as Google Autocomplete, discussed in chapter three. The social and economic impact of Google as a company, as well as its associations with other institutions, such as Facebook, are detailed in the context of digital capitalism in chapter five.

2.0 Literature Review

Search engines have been studied in various disciplines, for example, information retrieval, computer science, law, and new media, yet much of this work remains fixed within disciplinary boundaries. The approach of this thesis is to draw on work from a number of areas in order to link a technical understanding of how search engines function with a wider cultural and philosophical context. In particular, this thesis draws on critical theory in order to attend to the convergence of language, programming, and culture on a global scale.

The existing literature can be divided into a number of different areas, which bridge academic and non-academic study: academic computer science, Search Engine Optimisation (SEO) reports, business texts, legal and political approaches and broader interdisciplinary approaches. The current state of the field is outlined below.

2.1 Computer Science

The underlying technical processes of how search engines function are outlined in a number of texts. The original 1998 outline for Google, written by the founders Sergey Brin and Lawrence Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine” as well as the later 1999 paper co-authored with their Stanford professors, Rajeev Motwani and Terry Winograd (Page et al.), still provides a representative account of Google’s basic approach to search engine design. Further technical developments are outlined in *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW* (Belew), *Understanding Search Engines:*

Mathematical Modeling and Text Retrieval (Berry and Browne), and *Google's PageRank and Beyond: The Science of Search Engine Rankings* (Langville and Meyer). These texts are highly technical and written for a computer science audience; they provide a useful background but, as this thesis is focused on social and philosophical implications, I have chosen to limit the mathematical and programming details.

2.2 Search Engine Optimisation (SEO)

SEO is a large and growing industry and field of study; practitioners are employed by businesses or individuals to improve webpage rankings and operate in a grey-area in regard to Google. Due to Google’s secrecy about their algorithms, at its worst SEO is digital con-artistry but at its best the industry provides a rigorous interrogation of the day-to-day changes to Google's products as well as overall trends regarding search engines. Danny Sullivan, who is cited throughout this thesis, first established the industry in 1996 with his blog *A Webmaster's Guide To Search Engines* and went on to found the websites *Search Engine Watch* and *Search Engine Land*. Although the field focuses on the particular changes and practical advice, their overall reports provide useful technical insight into the ways that search engines have changed from the late 1990s until the present.

2.3 Business

There are a large number of books that address Google from a purely business perspective, such as *What Would Google Do?* (Jarvis), *I'm Feeling Lucky: The Confessions of Google Employee Number 59* (Edwards), and *The Google Guys: Inside the Brilliant Minds of Google Founders Larry Page and Sergey Brin* (Brandt), however, in the context of this thesis most of these texts can be ignored. Three books that balance the evolution of Google's business practices with insightful commentary on the changing context of search are John Battelle's *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, which, published in 2005, can be considered the first full-length study of search engines outside of computer science, Ken Auletta's *Googled: The End of the World As We Know It* (2009), and Steven Levy's *In The Plex: How Google Thinks, Works, and Shapes Our Lives* (2011). These are cited throughout as they contain interviews with Google engineers and provide a context for historical changes within the search engine industry.

2.4 Legal and Political

Siva Vaidhyanathan's *The Googlization of Everything: (And Why We Should Worry)*, published in 2012, was the first academic book-length study written in English to address the legal repercussions and social implications of Google's global dominance. The following year the Italian writers' collective Ippolita published a revised English version of their 2007 text *The Dark Side of Google*, which represents a more speculative, political, and overtly critical stance than Vaidhyanathan.

Google's CEO Eric Schmidt and Director of Ideas Jared Cohen's *The New Digital Age: Reshaping the Future of People, Nations and Business* was also published in 2013 and is a sourcebook for the idealism and techno-deterministic attitudes prevalent within the company. Julian Assange, one of the individuals interviewed for Schmidt and Cohen's book, presents his version of events in *When Google Met WikiLeaks.* Google's relationship with wider modes of capitalism is the focus of this thesis' final chapter; however, the political implications and potential alliances with the United States' government, including the National Security Agency's Prism program, although important, are not directly addressed in this thesis.

2.5 Edited Collections

A small number of edited collections and journal special issues focused on search engines have been published in the last decade. The key texts are as follows, *The Social, Political, Economic, and Cultural Dimensions of Search Engines* (Hargittai ed.), *Die Macht der Suchmaschinen / The Power of Search Engines* (Machill and Beiler eds.), *Deep Search: The Politics of Search beyond Google* (Becker and Stalder eds.), *Web Search: Multidisciplinary Perspectives* (Spink and Zimmer eds.), *Web Search Engine Research* (Lewandowski ed.), and *Society of the Query Reader: Reflections on Web Search* (König and Rasch eds.). These collect together many of the useful articles on search engines, referred to throughout this thesis. A recurring shortcoming, however, is that many of the articles provide isolated approaches or use search engines to provide evidence for a different issue, rather than focusing on them directly.

2.6 Interdisciplinary Studies

There are two book-length interdisciplinary studies of search engines that provide models that this thesis follows: Alexander Halavais' *Search Engines Society* (2008), provides a technologically focused study that places detailed analysis within a wider cultural and social background. Ken Hillis, Michael Petit, and Kylie Jarrett's *Google and the Culture of Search* (2012) is wide-ranging in its theoretical and philosophical sources and provides the context for chapter two. Both texts draw from a range of critical theory, philosophy, new media theory and other areas; while both have arguments or analyses that I argue against, their attempts to pull back focus and view search engines in a broader context is something that this thesis aims to emulate throughout.

3.0 Chapter Outline

The following chapters are structured in a broadly chronological way around two scales: the historical and the step-by-step actions of using a search engine. Chapter one addresses the formulation of search engine queries while outlining the early years of search engine development, from 1995 onwards. Chapter two focuses on the use of search engines as memory aids and establishes a much longer pre-digital history that stretches back to antiquity. Chapter three attends to the Autocomplete function that directs users as they type their query, in doing so reflecting on the changes established by Google from 2004 onwards. Chapter four considers the content of Google's search results and investigates the various criteria used to provide results by analysing the results for a set of queries searched in 2015. Chapter five addresses the global influence of Google from an economic standpoint and highlights a range of its effects, including the rise of fake news during the 2016 US Presidential election. Further detail regarding the methodological approaches and content of the following chapters is provided below.

Chapter one compares search engine queries to traditional questions. The chapter draws from information retrieval and computer science research to provide a technical framework that is brought into contact with philosophy and critical theory, including the work of Plato and Hans-Georg Gadamer. The chapter also surveys research on translation log analysis, the findings of which have not been addressed outside of information retrieval, to provide an insight into the search patterns of users between 1997 and 2006.

Chapter two investigates search engines as memory aids, deploying a history of memory and exploring practices within oral cultures and mnemonic techniques

such as the *ars memoria*. This places search engines within a longer historical context while drawing on contemporary insights from the philosophy and science of cognition. In particular, this chapter surveys psychology studies regarding the interdependence of memory on external aids, termed transactive memory, as a way of tying historical examples to current experimental findings that address the possible consequences of ubiquitous access to search engines.

Chapter three focuses on Google's Autocomplete functionality, which I argue represents a wider cultural shift for all search engines and other digital interfaces from 2004 onwards. Autocomplete is the most discussed search engine feature in popular media and the press, therefore, this chapter begins by addressing this widespread attention. However, I argue that, despite such interest, Autocomplete is widely misunderstood and by outlining the tool in detail this chapter shows that it produces some of the most complex issues within the study of search engines. In addition, Autocomplete has received almost no academic attention, therefore, this chapter opens up further questions for future research.

Chapter four addresses Google's search engine results, outlining them from a technical perspective and exploring the social implications of tailoring results for different contexts. This chapter also addresses the notion of *relevance*, a term used frequently by Google and search engine technologists; although this term is mentioned throughout the academic literature on search engines, no author has addressed its usage in detail. I explore two different readings of the term, which I argue are each employed with different epistemological foundations. The final part of this chapter presents an original study, carried out in 2015, that documents how different contextual signals, such as location, language, and phrasing of queries, change the kind of search results provided by Google. The same study was then

carried out in 2017, which provides an opportunity to analyse the changing nature of Google's results. In doing so, this chapter addresses the methodological challenges of studying algorithmically generated results that affect a wide range of disciplinary approaches. This chapter also raises questions regarding the globalising impact of digital institutions, such as Google.

Finally, chapter five explores how Google's method of generating revenue, through advertising, has a social impact on the web as a whole, particularly when considered through the lens of contemporary Post-Fordist accounts of capitalism. This chapter considers Google's effects on diminishing linguistic diversity online as well as the way in which the incentives of digital advertising incentivised the rise of fake news during the 2016 US presidential election.

Throughout, this thesis develops a framework for attending to algorithmic cultures and outlines the specific influence that Google's search engine has had on the web and continues to have on a global scale.

4.0 Challenges

There are a number of challenges regarding the study of search engines. Many of the objects of study, such as algorithms and datasets, are proprietary information, which companies do not release for various reasons, including fear of competitors adopting their methods or users gaming their ranking system. As discussed above, there is no independent index of the web; this means that researchers do not know the size of the web or what webpages are not returned by search engines or relegated to the unknown through low rankings. Google, Bing and other search engines admit to a certain degree of censorship, in alignment with local and international law, but without an independent index there is no way of finding out whether their censorship goes further than they claim. Google makes around “500-600” (Hargittai, “The Social, Political…” 772) changes to their algorithms a year and do not offer an historical record of search results; in addition, search results differ between individuals and across various contexts depending on a large number of factors, many of which are also kept secret. As Eszter Hargittai argues, this “poses significant challenges for the replication of search results, which is a basic tenet of scientific investigation” (772). Google’s expanding use of machine learning also introduces situations in which various emergent technosocial arrangements are unintelligible even to the engineers responsible. These, and other issues, are addressed throughout this thesis and developed more fully in chapter four, specifically in regard to analysing the results of search engines. These challenges can be frustrating, both for author and reader, but highlighting, rather than avoiding, them is central for an honest reflection of contemporary digital culture. This thesis

aims to draw together tactics used by other scholars, while also introducing new ways of engaging with algorithmic cultures.

5.0 Notes on Notation and Examples Used

Throughout, I have followed the standard practice, within search engine studies, of using [square brackets] to signify when a word or a phrase is being referred to as a search engine query. Note that [sic] applies to them throughout, as when using my own examples and those from other researchers, these reflect direct queries and therefore often lack capitalisation, contain slang, and are often grammatically incorrect.

There are times when screenshots of Google searches are used as evidence or examples. All of these searches were performed using a Virtual Private Network (VPN) in order to anonymise search behaviour, prevent results from being personalised, and to control other kinds of information sent to Google regarding the search. At times, in particular in chapter four, the VPN is used to adopt an Internet Protocol (IP) address from a server in a different geographical location, in order to compare the searches from one country to another.

Chapter One:

From Questions to Queries: The Changing Role of Questioning Through Search Engines

Introduction

This chapter's focus is twofold: first, a technical description of how search engines function is outlined; second, this chapter places web search queries within a broader philosophical context regarding the nature of questions. The initial section explores the computational techniques that underpin how search engines function and demonstrates how even the most basic processes have deeply embedded social and cultural attitudes. Doing so establishes a major theme of this thesis that technology is never neutral or apolitical.

The technical aspects of how any web search engine functions can be split into three parts: crawling and scraping, indexing and ranking, and queries and results. In the course of outlining these three general processes Google's particular innovations are highlighted, as their impact shaped industry norms. Google's methods set their search engine apart from their competitors and led to their large-scale influence, in terms of market share and technical changes implemented by other search engines in response. In doing so, this analysis demonstrates why Google is of primary focus throughout this thesis. Although this chapter provides insight into the early technical developments of search engines, it does not attempt

to provide a detailed history of the rise and fall of particular technology companies. Elizabeth Van Couvering's article "The History of the Internet Search Engine: Navigational Media and the Traffic Commodity" provides the most thorough chronology of the search engine industry between 1994 and 2006. An alternative history is provided by John Battelle in *The Search*. In the chapter "Search Before Google" Battelle includes a slightly wider range of examples, such as the Archie tool implemented in 1990, although its designation as a web search engine is, like most firsts, widely debated.

The second key focus of this chapter is to explore how search engine queries are related to traditional modes of asking questions. The interface of search engines requires users to provide language in order to gain access to information, as is the case in other modes of enquiry; however, the technology behind how search engines function means that this process treats language in a radically different way compared to the conventional discourse of asking questions. Although the technologies underlying how a search engine query is treated have become increasingly complex, the fundamental difference between questions and queries remains: queries are an automated search for the words and phrases submitted. Even if users frame their queries as questions, the way in which queries are interpreted means that linguistic inputs establish the kinds of results provided, in that they mirror the user's language. In order to ground such claims, this chapter draws on a range of studies from the discipline of information retrieval (IR) that analyse actual search queries of users, between 1997 and 2006. An overview of this literature provides an empirical body of evidence that is used throughout the thesis.

After establishing a technical outline of search engine queries, these insights are placed into a broader philosophical context, including Plato's discussions

concerning questions and Hans-Georg Gadamer's hermeneutical model of enquiry. The interdisciplinary nature of bringing together work from the social sciences, philosophy and critical theory is this chapter's original contribution, as the statistical conclusions regarding search engines have not been widely cited or analysed outside of their original discipline. This chapter also establishes the foundation for chapter two, which addresses the longer history of various technosocial memory practices in the context of current debates regarding the effect of search engines on memory.

1.0 The Three Steps of Search Engines

1.1 Step One: Crawling and Scraping

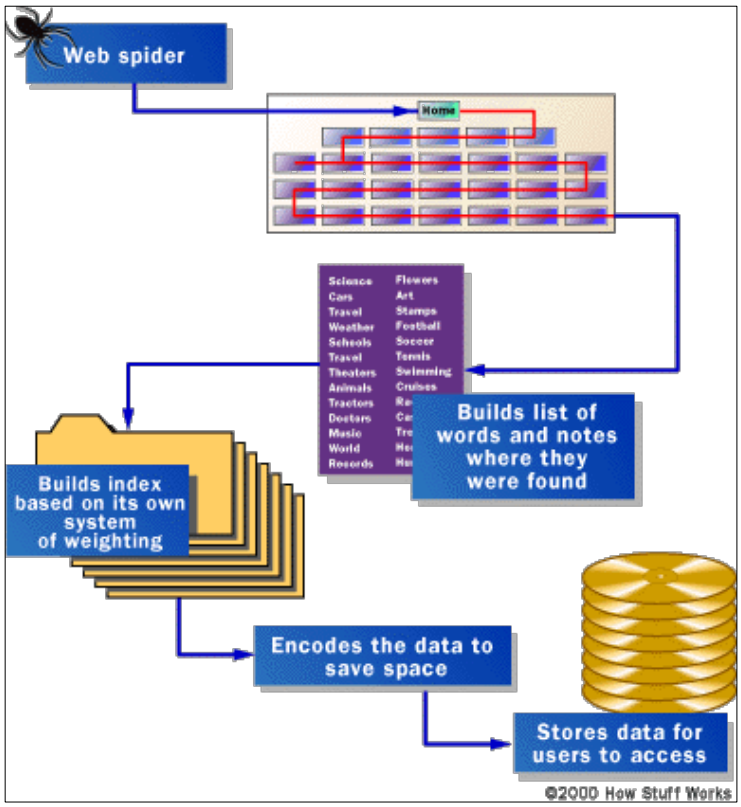


Figure 1. A diagram outlining the first two steps: Crawling and scraping, and indexing and ranking. Taken from Curt Franklin “How Search Engines Work” *How Stuff Works*.

At the most basic level, search engines function like an index in the back matter of a book, however, unlike an index’s usual subject specificity, search engines record every word they find on the web. When an individual uses a search engine they are not searching the web directly, instead, they are searching a particular index of the web. Each search engine builds its own index through an automated software application known variously as a “crawler, spider, web robot, or bot” (Halavais 14).

These spiders crawl the web by carrying out two simple tasks (see fig. 1). First, a webpage will be chosen at random and the spider records all the words that appear on that page (ignoring very common words such as ‘the’ and ‘and’). Next, the spider looks for any hyperlinks on that page and follows each of them in turn, repeating the process of recording the words used on each page. In doing so, the results for all spiders can be aggregated and an index of words and locations can be built up. A search engine index differs from that of a book in two ways; first, they cover every word the spiders find, due to the open-ended nature of the web; second, as search engines advanced, other kinds of information about the relationships between words and documents were stored in the index. The first basic search engines (WebCrawler, Lycos, Infoseek, Open Text, Excite, AltaVista, Inktomi; see Van Couvering “The History” 183) worked as a blunt tool whereby a user’s query – their linguistic input to search engine – represented a request interpreted as “show me all the instances where this word can be found”. Because indexes are automated rather than constructed by a human editor, as they would be for a book, the nuance of contextual significance is lost.[1] Although search engines have grown in complexity, this structure, whereby search queries represent language a user hopes to find within an online page, remains their foundational logic. Attending to the systems that support such a logic, demonstrate that although queries are part of a wider tradition of enquiry, their technological characteristics make them distinct from previous non-computational modes of discourse.

Whereas the processes of indexing and ranking have evolved considerably, their development shaping the successes and failures of various search engines, this initial step of crawling the web has not changed significantly. As Alexander Halavais

[1] There were attempts at editorial search engines, indexed by teams of individuals, such as Magellan, but such a method could not scale to the size of the web, even in its earliest days.

outlines, “the crawler is a relatively simple piece of technology [and] it has not evolved as much as other parts of the search engine” (15). Although crawling itself is simple, other factors cause its implementation to be uneven and complex. If users do not want their page to be crawled and added to an index, webmasters may employ a Robots Exclusion Protocol, also known as robots.txt, which stops spiders from cataloguing explicitly exempted pages. Consequently, although theoretically any organisation crawling the web with spiders should build up an identical index of the web, a robots.txt file can specify which particular spiders are denied and which can crawl their page. Statistical analysis by Yang Sun found that “popular search engines and information portals such as Google, Yahoo and MSN are generally favored by most of the websites we sampled” (83); in particular, “Google is always the most favored robot” (28). Therefore, although the technology deployed by different institutions is the same, Google captures more of the web. Therefore, a feedback loop regarding institutional trust emerges: Sun’s results demonstrate that smaller or lesser known institutions are subject to a larger number of robots.txt denials compared to more established institutions leading to a bias whereby the “rich get richer” (29).

There are some specific consequences of the technology behind scraping remaining relatively unchanged. The main consequence is that it has led to the separation between the surface web (i.e. pages that can be scraped by spiders) and the deep web, which constitutes any pages that cannot be captured for various reasons. The deep web should not be confused with the dark web. The dark web is a subdomain of the deep web and is often used for criminal activity such as the Silk Road, however, the two are regularly conflated by media reports, which often gives way to a misunderstanding that pages not indexed by search engines are

objectionable or undesirable.[2] There are many different kinds of pages that cannot be found using spiders, a large proportion of them are pages that do not have any direct incoming hyperlinks, such as databases or other sites that require different kinds of inputs to generate their pages. Because of this "information contained in these hundred thousands [sic] of repositories is […] hidden from users" (Shestakov and Salakoski 780). Therefore, any estimate of the size of the deep web represents little more than guesswork. There has been much research into ways to overcome this information gulf,[3] however, the dominant way that pages from the deep web can become part of the surface web is through webmasters providing an XML file called a sitemap, developed and released by Google in 2005.[4] Sitemaps allow webmasters to indicate the architecture of pages they wish to be indexed, including those lacking incoming hyperlinks. In a similar situation to the robot.txt exclusion bias detailed above, sitemaps are submitted to specific search engines so are likely to follow similar biases. Because of these challenges, which are mostly overcome through the actions of webmasters, the "index knows a lot about the information from the point of view of the providers who create the topology of the network" (Stalder and Mayer 98). Therefore, social attitudes, market dominance and economic incentives have a direct bearing on the most fundamental parts of how search engines function. As will be seen throughout this thesis, it is misleading to describe even the simplest technologies as neutral, or to suppose that technologies can ever exist outside of their cultural context. The interplay between web-crawling spiders, robots.txt,

---

[2] Andy Greenberg's "Hacker Lexicon" provides more information about the distinction and a number of examples where various established media institutions have reported wildly inaccurate stories due to conflating the two terms, which Greenberg implies may well be a wilful conflation.

[3] One method uses a free software module called mod_oai that can expose the contents of databases such as digital libraries. This module was built by Michael Nelson and documented in Nelson et al. "mod_oai: An Apache Module for Metadata Harvesting" and "Efficient, Automatic Web Resource Harvesting".

[4] See Danny Sullivan's 2005 interview with Google's engineering director Shiva Shivakumar for greater detail.

sitemaps, web users, and webmasters exemplifies Bruno Latour's statement, in his 1990 essay "Technology Is Society Made Durable", that "we might call technology the moment when social assemblages gain stability by aligning actors and observers. Society and technology are not two ontologically distinct entities but more like phases of the same essential action" (129). The technological interactions have attitudes embedded within them; this means that by treating a single search engine, Google for example, as the most reliable or effective, such an affirmation becomes a self-fulfilling prophecy. In addition, as particular cultural attitudes are stabilised they can become treated as neutral or apolitical. Evidence for Latour's assertion regarding the enmeshed nature of the technosocial is presented throughout this thesis, first in chapter two, which highlights the historical continuity of such a claim and then in chapter three, which presents the case that Google's algorithms directly represent particular cultural values.

1.2 Step Two: Indexing and Ranking

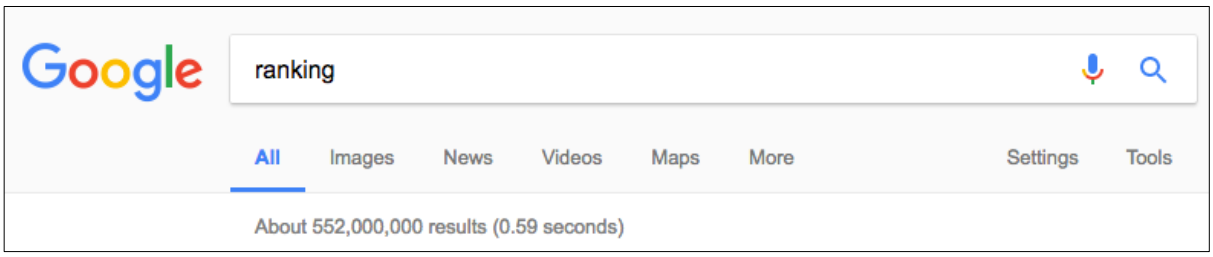


Figure 2. A search for the query [ranking], which results in "About 552,000,000 results".

The process of crawling the web with spiders results in a large and crude index of words that might become key terms in a future search engine query. The scale of the

web means that listing all pages that contain a search term in an arbitrary order, such as alphabetically, or ranking them based on sheer quantity of a word would not be considered practical (see fig. 2). In addition, web content has a lower barrier to entry than print publication, which leads to a broader range of both subject areas and quality. Therefore, the history of which search engines succeeded and which failed is largely based on the effectiveness of indexing and ranking pages. For example, fig. 3 compares two potential webpages that would be listed by any search engine in response to searching the query [Shakespeare].

Figure 3. Comparison between two webpage results for the query [Shakespeare]: a high-ranking BBC article written by Martin Wiggins (left), and a hypothetical low-quality webpage result (right).

The ranking of search results serves to prioritise high-quality results, such as the BBC article on the left, while deprioritising low quality or spam pages. At the heart of this process is an attempt to automate complex, hermeneutical human judgements. Different search engines have prioritised different criteria for automating such

judgements, including measuring linguistic importance, interpreting design decisions, and aggregating widespread human sentiment. Building a program to make those kinds of judgements is difficult. So, instead of implementing a system that attempts to interpret each page from scratch, search engines were designed to measure the human sentiments latent in the language used, the kind of unavoidable design decisions implicit in the code, and the existing behaviour of users.

The overall history of the search engine market has been well-documented by Van Couvering and Battelle, therefore it is not necessary to repeat their work in detail; however, there are two examples within that wider narrative that shed light on particular technological perspectives that had a large impact on the evolution of the web and the wider online information ecology. One of the first search engines to become majorly successful was Alta Vista, launched in 1995, two years before Google. The focus of AltaVista's ranking emphasised webpage design, in particular the hard-coded context of language, such as where and how words were displayed on a webpage and their proximity to other related words. This was accomplished by placing high importance on key elements of the computer code used to format webpages: Hypertext Markup Language (HTML) metawords. This used the existing format of pages to provide an indication of word importance. For example, if a word is coded as a title or header of a page, then this page would be ranked higher than another page that contained that search term in the body of the page. Fig. 3 provides an example of a header tag emphasising the word Shakespeare and how this might be used, in the context of fig. 2, to rank this page higher than others for the query [Shakespeare].

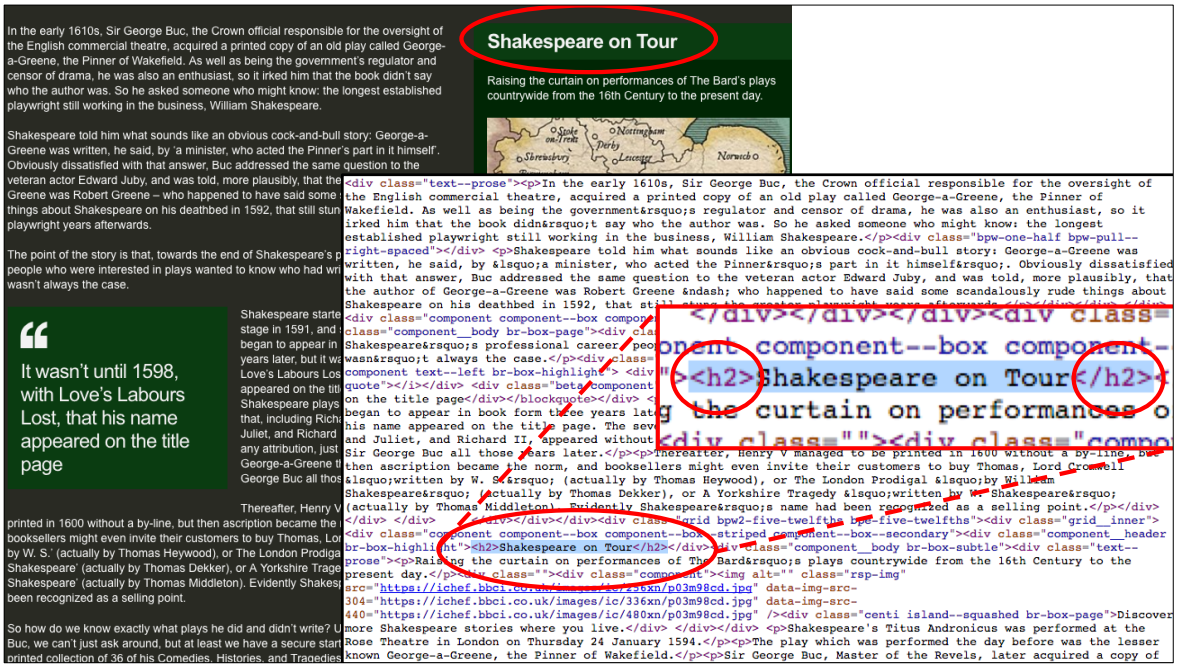


Figure 3. This image demonstrates how design decisions coded in HTML, such as including subheaders (<h2>), can be read by search engines and used to rank search engine results. Drawn by the author.

In addition, AltaVista prioritised the distance between various search terms. This allowed the recalling of information from the index to be based on the principle that, “pages where the query words occur near each other are more likely to be relevant than other pages where the query words are far apart” (MacCormick 18). John MacCormick provides an example of a search for [malaria cause] (18), which would cause a search engine to search its index for the two words separately: malaria and cause. When deciding the ranking of each result, MacCormick invents two possible results: one with the phrase “the cause of malaria…” (19), and the other which starts “the cause of our expedition was not helped when…” (20) and which mentions malaria on the same page but in a later section of the writing (see fig. 4).

Figure 4. An illustration (drawn by the author) to demonstrate MacCormick's example of how word proximity can indicate probabilities of semantic relevance.

The metaword algorithm would, therefore, present the first result as more relevant than the second and would present another page with both malaria and cause in the title as more relevant still. This system of ranking worked better than AltaVista's competitors when it was launched and is still one of the many metrics taken into account by contemporary search engines; however, it was not long before AltaVista was overtaken by Google, which pioneered a different way of prioritising and ranking pages. Google's primary method is still in use and not only has it remained dominant, but it has shaped the overall architecture of the web, which has reinforced its influence.

Although there are a number of different reasons why Google became the most dominant online search engine, there was a specific technological difference between its search engine and other competitors. The founders of Google, Sergey Brin and Lawrence Page proposed the idea of increasing the importance placed on

the hyperlinks between webpages and reducing the significance of the HTML layout of each page. [5] The evaluation of links would work in a way similar to how academic citation, in aggregate, conveys that importance of a source within a particular field or discipline. The more links to a page, the more important the community had deemed that page. However, as they describe, there are

> significant differences between web pages and academic publications. Unlike academic papers which are scrupulously reviewed, web pages proliferate free of quality control or publishing costs. With a simple program, huge numbers of pages can be created easily, artificially inflating citation counts […] Further, academic papers are well defined units of work, roughly similar in quality and number of citations, as well as in their purpose – to extend the body of knowledge. Web pages vary on a much wider scale than academic papers in quality, usage, citations, and length. A random archived message posting asking an obscure question about an IBM computer is very different from the IBM home page. (Page et al. 1-2)

Nonetheless the academic citation model would still be useful if the right measures were accounted for; in particular, the aspect that citations from more important or influential sources should be weighted to a higher degree than unimportant sources. Therefore, a calculation was required to factor flows of influence measured by an algorithm they named PageRank. Therefore, pages with a high number of incoming links or links from important sources would gain a score that would then be carried over to any pages subsequently *linked to* by this important page. Through this

[5] Documented in Brin and Page “The Anatomy of a Large-Scale Hypertextual Web Search Engine” (1998) and Page et al. “The PageRank Citation Ranking: Bringing Order to the Web” (1999).

method Brin and Page created maps of authority that they argued "allow rapid calculation of a Web page's 'PageRank', an objective measure of its citation importance that corresponds well with people's subjective idea of importance" (109). Hyperlinks do not necessarily provide value judgements, for instance, when a webpage author links to a page because it wants to flag it as bad, there is no criterion which categorises links as negative or positive. In the words of MacCormick, however, "it turns out that, in practice, hyperlinks are more often recommendations than criticisms" (27).

Google's PageRank algorithm functions by co-opting an aggregate of intention, as Jon Kleinberg argues: "hyperlinks encode a considerable amount of latent human judgment, and we claim that this type of judgment is precisely what is needed to formulate a notion of authority" (2). Each hyperlink has a different level of value; for example, an outgoing hyperlink from a page like whitehouse.gov carries more value than an outgoing hyperlink from the page of an amateur blogger. In turn, because these values are relational, the page that the whitehouse.gov hyperlinked to must also be given a high value. If the whitehouse.gov links directly to the amateur blogger's page, then any outgoing hyperlinks from that page now carry an additional value derived from the previous hyperlink. Hyperlink judgements are all weighted in a way that is exclusively dependent on other nodes in the network. In order to prevent feedback loops the algorithm employs a Markov Chain, which is a method of measuring these values as probabilities.[6] However, because the values are reliant on each other, existing relationships of influence become reinforced. For example, even if pages from the deep web are indexed through mod_oai or sitemap files, these pages cannot be provided with a rank based on anything other than internal

[6] For an in-depth mathematical explanation of the PageRank and its use of a Markov Chain, see pages 84-88 of Berry and Brown *Understanding Search Engines: Mathematical Modeling and Text Retrieval*.

links. Therefore, PageRank's logic prioritises existing hierarchies and thus trends towards a more static informational landscape.

Although there are many ways in which Google's results have become mutable in a range of contexts, through localisation, personalisation, and time-dependent variables, which will be discussed in further chapters, Google's original design was focused around reflecting an existing order. Page et al. in their 1998 paper describe their initial aim in the following way:

> The importance of a Web page is an inherently subjective matter, which depends on the readers [sic] interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. [PageRank is] a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them. (1)

Unlike earlier uses of HTML metaword coding which measured the value of words, their proximity, and the formatting of pages independently of one another, the emphasis of Google's ranking algorithm placed the greatest importance on networks between pages. The coding and format of each page was still taken into account, but played a significantly lesser role in establishing the overall information ecology on the web. This codified outlook directed the landscape of the contemporary web to be dominated by large or well-known institutions by establishing what could and could not be discovered through a web search. Google's influence, therefore, developed in concert with existing distributions of power and influence. Therefore, PageRank should be considered as a way of harnessing existing relationships, a type of

practice which, to borrow from Andrew Pickering's concept of the mangle of practice, "should be seen as a process of modelling, of the creative extension of existing cultural elements" shaped by technological and social contingencies. (Pickering and Stephanides 140). Modelling is shaped by "resistances" (Pickering 23), which leads to "accommodations" (23) as new behaviour is set in place. In 1998, Google established a logic that aligned quality with citation and provided a model for how the web ought to look in a way that became recursively influential. As will be developed throughout this thesis, Google's original model has both established and adapted to, the changing nature of the web.

1.3 Step Three: Query and Results

The final stage of the search engine process is the front-end that users interact with. The search box, in particular Google's minimal surrounding layout, has become a ubiquitous part of the web. The early web portals that predate search engines rearticulated an old media aesthetic of a newspaper front page or Teletext offering suggestions and directions to a user. Query formation by users will be addressed later in this chapter and in chapter three; search engine results are discussed in detail in chapter four. For the moment, they receive a brief explanation to complete the basic description of how search engines function.

The blank search box of a search engine can take any kind of query: originally, Google allowed any combination of characters up to a maximum limit of thirty-two words, however, the evolution of search engines now enables searching through spoken word and even using pictures. The way in which search engines

interpret a query has changed significantly over the years; originally it would look for exact word matches in a document from its index responding to the HTML tags, as explained above, combined with its algorithm to judge citation value. Google also offered a selection of operators to narrow the search, the Boolean AND, OR and NOT as well as other characters that explicitly told the search engine to include synonyms (~ (tilde)), include wildcard terms (*), look only for exact phrasing (“ ”) etc. Although these operators have been gradually phased out and replaced by automated systems outside of a user’s control, [7] the main underlying logic of formulating a search engine query is entering words that the user hopes to find on a page. In a very literal sense, search queries embody Wittgenstein’s statement from his *Tractatus Logico-Philosophicus* that “*The limits of my language* mean the limits of my world” (68 emphasis in original). The linguistic nuances of each query, including variations of vocabulary, tone, grammar and phrasing, shape the kinds of webpages that are provided through search results. In addition, every time a query is searched the results are ranked *sui generis* according to a range of criteria, also referred to as signals, used to predict that specific user’s intentions and outlook. Google claim to use around 200 individual signals such as information about the user: their location, the time of day, and their browsing history; as well as information about the pages found (how long a site domain has been registered, the location of the server on which that page is stored, how long it will take to load).[8] This means that search

[7] Interface simplicity has always been a priority for Google, but there are other potential reasons for minimising the range of search setting available to a user. For example, providing fewer option for users limits the variety of queries, thus enhancing Google’s purview over search results. For further commentary, see Barry Schwartz’s “Google Drops Another Search Operator: Tilde For Synonyms”, which also includes links to five other reports of search functionality dropped by Google.

[8] The number of 200 has been used by Google (see Cutts) as well as the academic community (see Levy *In the Plex* 49). See Brian Dean’s “Google’s 200” for a compilation of the existing evidence concerning the nature of these signals. Eric Schmidt, Executive Chairman of Alphabet, considers these factors to be “a business secret of Google” (see Sullivan “Schmidt”). The number may well be arbitrary, as in 2010 after Bing announced they use 1,000 signals, Google responded that their 200

engine results can be different between individuals depending on a number of factors. The criteria involved in these 200 signals are complex, therefore, chapter four directly explores the kinds of signals that influence which results are provided for different contexts and the ways in which search results for controversial topics can differ radically as a consequence.

There are many differences between questions and queries, but the way in which queries must mimic the language of their desired answer is of key importance. Even though Google Search employs synonyms by default, the extent to which alternative words and phrasing are used to supply results is highly variable.[9] In Steven Baker's official Google blog post, "Helping Computers Understand Language", he uses the example of how, through using synonyms, a search for the query [pictures developed with coffee] is interpreted similarly to [photos developed with coffee] by drawing from results that contain the alternative wording. Fig. 5 shows the two searches, carried out in 2017 rather than 2010, with the shared results crossed out in red. There are some small differences but the linguistic replacement has not caused much variation in the results. This is in sharp contrast to, for example, words that have more complex associated uses. Fig. 6 compares the results for the queries [being a woman] and [being a female] in which only one result, "12 Things About Being a Woman That Women Won't Tell You" written by Caitlin Moran for *Esquire*, is consistent in the first page of results. The differences shown in fig. 6 are ripe for a cultural critique, regarding the relationship between gendered experience, linguistic patterns, and modes of communication that is

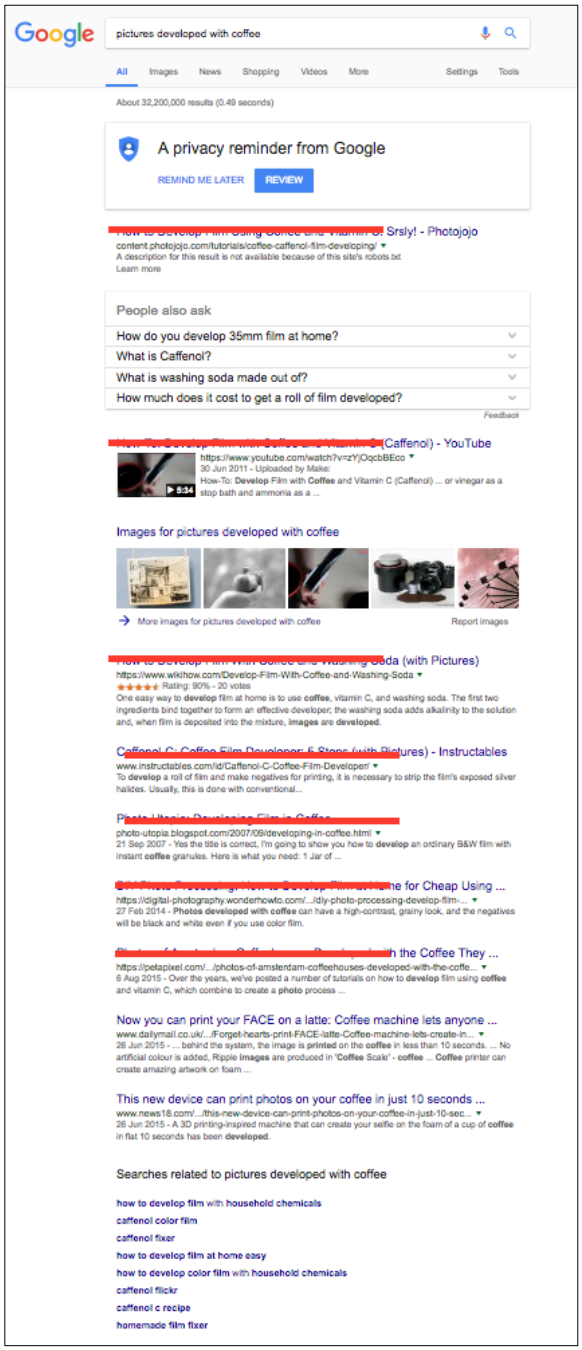
---

signals are each made up of 50 sub-signals, thus raising their total to 10,000 (see Sullivan "Dear Bing").

[9] Google explicitly acknowledged that they use synonyms by default in an official blog post in 2010, (see Steven Baker's "Helping computers"), but the practice is most likely much older, given that it relies on a particular patent filed by Baker and John Lamping in 2005 (see Baker and Lamping "Determining query term synonyms within query context").

beyond the scope of this chapter; chapter four addresses how similar linguistic shifts alter search results directly. Figs. 5 and 6 are included here in order to demonstrate that sometimes differences in vocabulary greatly influence search results and sometimes they do not. Like much of Google Search's interface, these variations are not explicitly communicated to users.

The explanation provided above outlined that queries are fundamentally different from questions. However, as shall be explored below, search engines facilitate a much wider range of online behaviours than even the broadest definitions of enquiry. What makes search engine interfaces somewhat unique is their lack of explicit framework. Web portals, in which the search bar is surrounded by curated information, dominated before the rise of Google and still exist today (see figs. 7-10). However, the search engine query bar at the centre of the pre-selected content still required the user to begin the process; even with the later implementation of Autocomplete, which will be discussed fully in chapter three, the drop-down list of suggestions still only appears after the user has initiated the process.



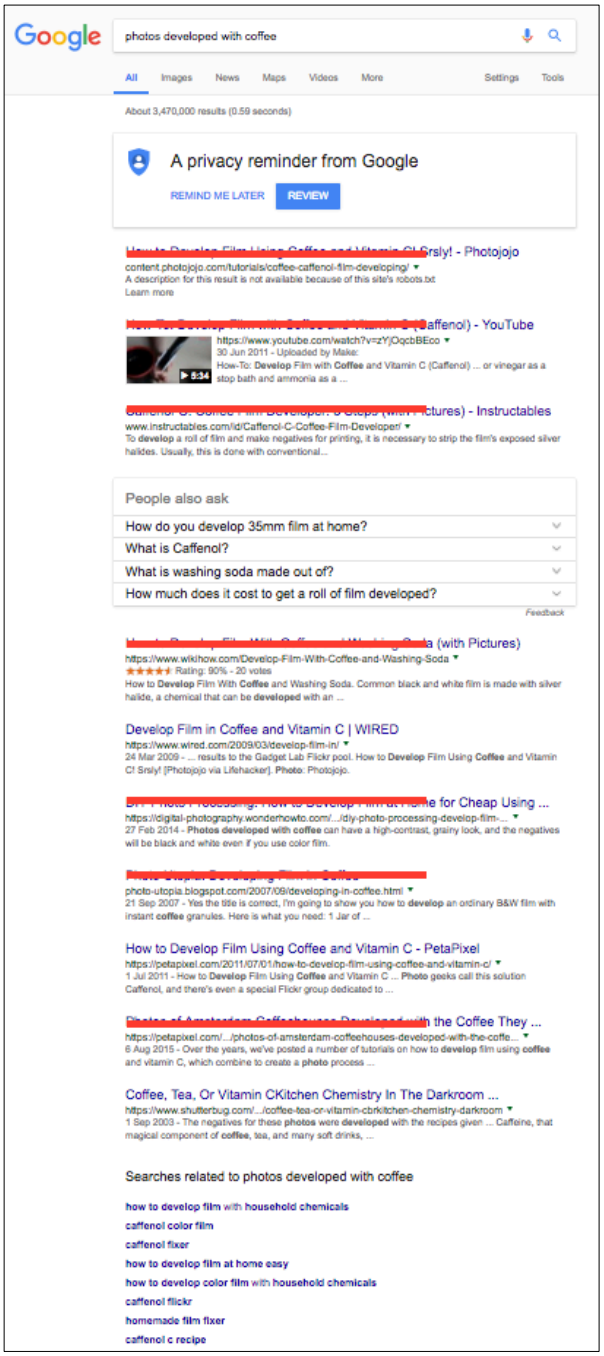


Figure 5. Two searches with slight linguistic variation in the query, suggested as an example in Baker's 2010 blog post: [pictures developed with coffee] (left), [photos developed with coffee] (right). Results that appear for both queries have been struck-through. Due to the use of synonyms the results are mostly the same. Search performed 10/10/2017.

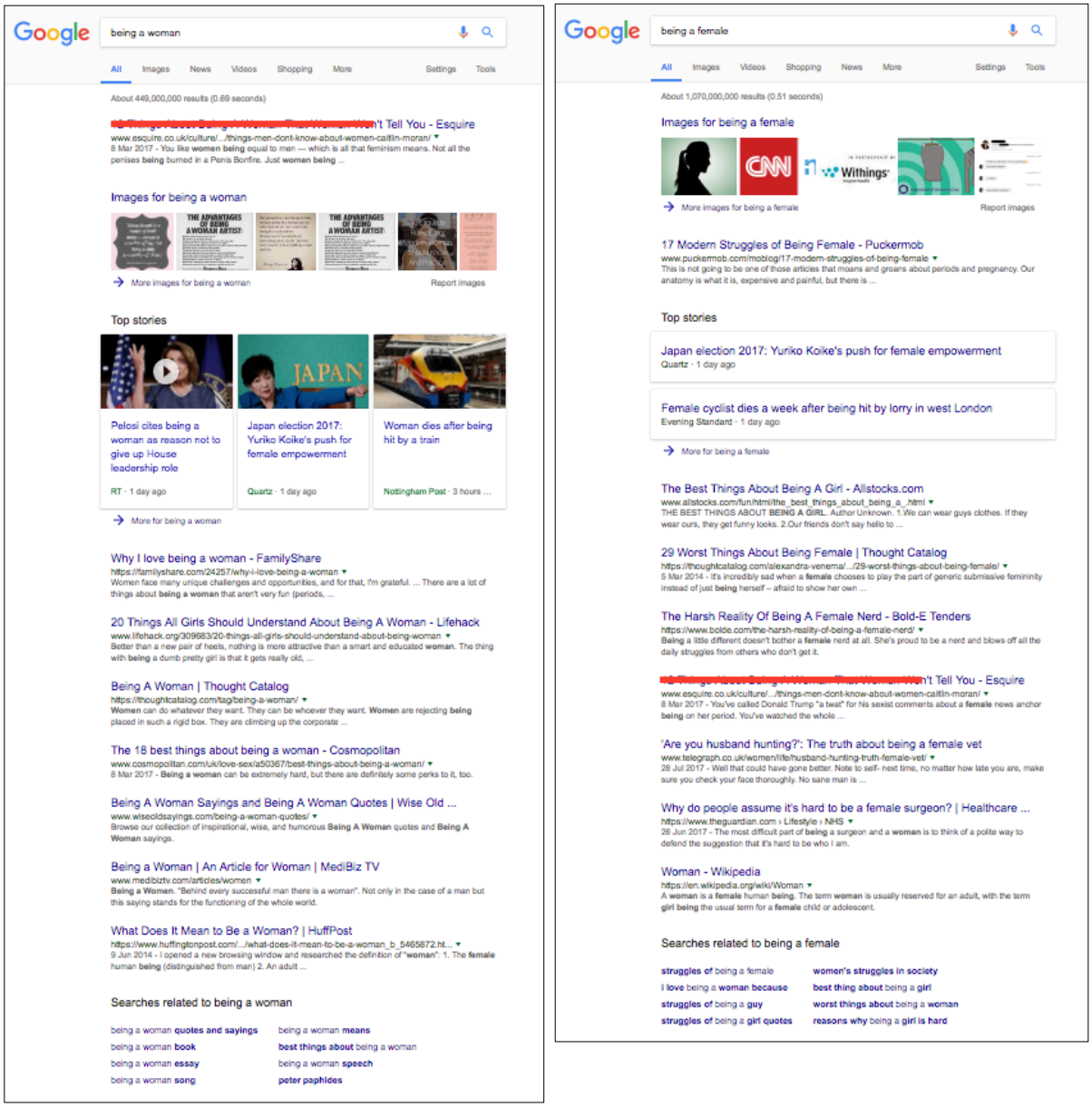
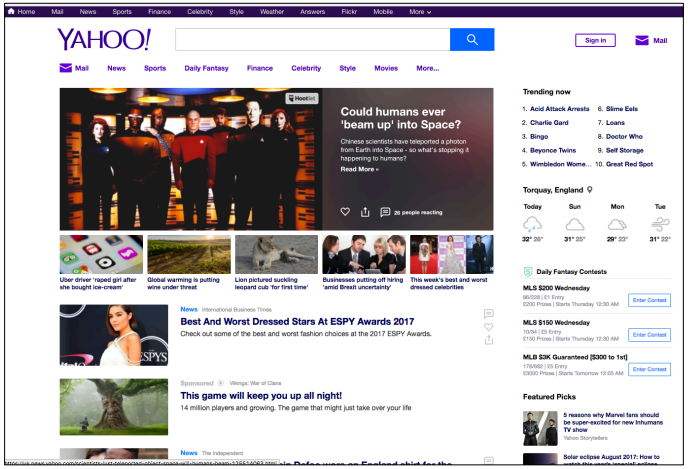
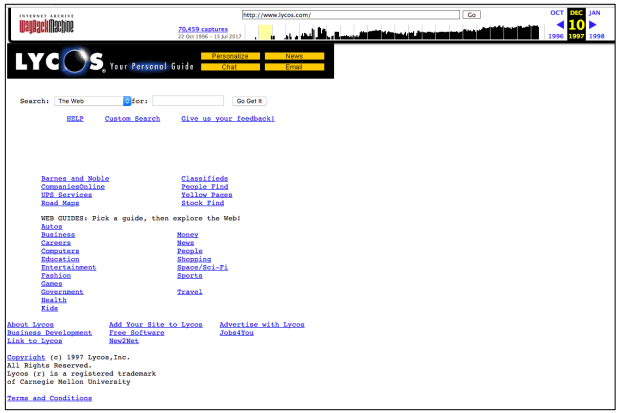
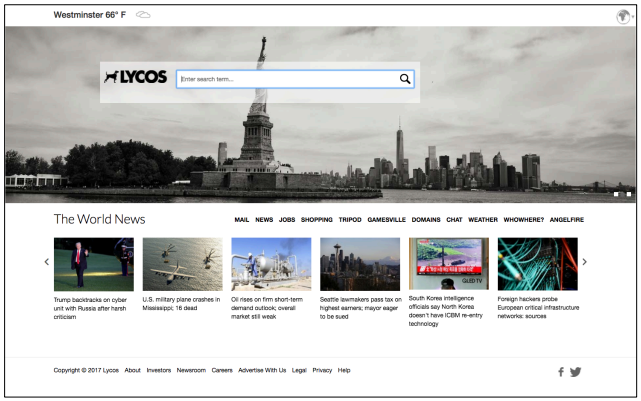


Figure 6. Two searches with slight linguistic variation in the query that show a much larger variation in search results. [being a woman] (left), [being a female] (right). Results that appear for both queries have been struck-through. Even though synonyms for the queries have been used, the results are very different. Search performed 10/10/2017.







From left to right. Figure 7. Yahoo! Web Portal, as of 17/10/1996, accessed via the Way Back Machine Figure 8. Yahoo! Web Portal, accessed 15/07/2017. Figure 9. Lycos Web Portal, as of 10/12/1997, accessed via the Way Back Machine. Figure 10. Lycos Web Portal, as of 15/07/2017.

In this way, Google stands apart from much of the history of information organisation. A survey of pre-digital information technologies – library categorisation systems, encyclopaedia, index cards – provides examples of infrastructure in which the information seeking behaviour might be prompted by a question, but which are organised according to pre-arranged, top-down hierarchies that do not incorporate question-asking behaviour into the system.[10] Google's interface makes no attempt to provide a detached perspective on the web. Most previous technologies allow browsing without intent: a user of a library using the Dewey Decimal System can walk to the shelves containing the 800 class books and browse all the literature that that library has; the owner of an encyclopaedia can read every entry that starts with A, if they so choose. That every use of Google requires some informational input on the part of the user represents a radical shift for technologies used for enquiry. The cultural and philosophical implications of such a shift are an important part of understanding Google Search, however, there is no existing literature that attends to the mandatory nature of queries in any real depth. Battelle coined the term "database of intentions" (2) in 2005 to describe Google's relationship to its users and the phrase has been taken up in subsequent studies (see Halavais 30, Duguid 16, Pariser 103, Hillis et al. 14, Jarrett "A Database of Intention?" 17, to name only a sample); this captures part of the interaction but has mainly been used in the context of privacy concerns. Hillis et al. put Google's dialogic nature in a religious context (171) as well as invoking a Zizekian Big Other (179-185) that leads their analysis into conflating the technical differences between questions and queries. Halavais notes that Google is fulfilling the role of gatekeepers such as journalists and teachers (105) but does not make explicit the part of this interaction that would be built around a

[10] See chapter two for a more in-depth assessment of the history of information technologies including examples that problematise this characterisation, such as oral culture and the *Ars Memoria*.

question and answer process. The only example of analysis that attends to the technological process of search engine queries, while relating this to a wider context of questioning is Boris Groys' "Google: Words Beyond Grammar". Groys' piece is theoretically ambitious in relating queries to the implications of Derridean Deconstruction. However, the piece is very short and, more significantly, presents some misunderstandings about the technical nature of Google's search engine. Therefore, the following section will introduce Groys' argument while providing some technical insights and a selection of quantitative studies that reinforce Groys' theoretical arguments, as well as some evidence that quashes some of his claims. After this technical basis has been established, I will then extend Groys' provocations by widening the historical narrative by introducing some of Plato's concerns. After doing so I then take the distinction between questions and queries a step further by introducing the work of Hans-Georg Gadamer, which I argue provides a useful theoretical perspective that widens the philosophical scope of questioning while keeping the technical aspects of search engine queries intact.

2.0 Google: Words Beyond Grammar

> Human life can be described as a prolonged dialogue with the world. Man interrogates the world and is interrogated by the world. This dialogue is regulated by the way in which we define the legitimate questions that we may address to the world or the world may address to us – and the way in which we can identify the relevant answers to these questions. (Groys 147)

Groys' piece opens with an emphasis on rules and the way that in order for questions to be asked they must be expressed in a legitimate form that responds to a given context. Search engines, on the other hand, do not require inputs in the form of questions. The way in which Google responds to queries is by providing links to "all the accessible contexts in which this word [or combination of words] occurs (148). This enables queries to cover a greater remit than traditional questions; however, without the logical, rhetorical or dialectical structures that characterise traditional questions, queries act in a very different manner. Groys sees the way that Google queries attend to language as an extension of the non-normative claims of Ferdinand de Saussure and Jacques Derrida in which the meaning of words is dictated by context and endlessly shifting chains of signification. Groys goes on to argue that

> Google, therefore, can be seen as an answer to deconstruction in at least two ways. On the one hand, Google is based on the same understanding of language as topological space, in which individual words follow their own trajectories - undermining any attempts to territorialize them in fixed,

> privileged, normative contexts and to ascribe to them normative meanings. On the other hand, Google is nonetheless based on the belief that these trajectories are finite, and so can be calculated and displayed. […] One can say that Google turns deconstruction from its head onto its feet by replacing a potentially infinite, but only imaginary, proliferation of contexts with a finite search engine. (150-151)

Here, Groys provides a fruitful frame of reference. Much of the existing literature that investigates search engines derives from an information retrieval (IR) disciplinary background and frames this philosophical shift only in terms of success and failure. For example, John Battelle's early study of search engines (published in 2005) compared the results for the search query [usher] of Yahoo and Google (237-239) in which, he "presume[s] the person typing that search in reality does want to know about the popular singer by the same name" (238). He then compares the way in which both search engines deal with results that are unrelated to Usher the singer, for example "The Fall of the House of Usher" and the communication disorder "usher syndrome". A full discussion of how various hangovers from IR have shaped search engine evaluation, particularly the elusive metric of relevance, takes place in chapter four. A range of social science techniques exist for numerically measuring success or relevance of search engine results.[11] "Search Engines Evaluation", written by Kumar et al. is a good example of such literature, the conclusion of which begins, "Lord Kelvin once said that when you can measure what you are speaking about and express it in numbers, you know something about it" (9). Such a simplification does not describe all IR research, but there is certainly a tendency within the field to tune-

[11] See Lewandowski "A Framework" for a literature review of "Retrieval Effectiveness Tests" and Jansen and Molina "The Effectiveness" for a literature review of "Web Search Engine Relevance Evaluation".

out richly incalculable moments, in any medium of discovery. Groys' deconstructive context helps foreground the shifting nature of meaning that the form of search engine queries accentuates. One particular overarching narrative in the last twenty years of search engines is the importance of context and the proliferation of vantage points, a topic which will be thoroughly outlined in chapter four.

As Groys' piece continues, he stretches his claims into conclusions about Google's effect on language that simply have no bearing on any available evidence. However, highlighting why his argument fails to reflect actual search engine technology provides a useful opportunity to further outline the differences between questions and queries. As "Google: Words Beyond Grammar" draws to a close Groys' introduces Martin Heidegger's notion, taken from his "Letter on 'Humanism'", of "language as a house of being – a house in which man dwells" (151), which image Groys uses to argue that Google Search has caused "the liberation of individual words from their syntactical arrangements [which] turns the house of language into a word cloud. Man becomes linguistically homeless" (152) and in turn, this has created a situation in which "all words are already recognized as 'metaphysically' free and equal" (155). Heidegger's original point in the "Letter on 'Humanism'" is about the impoverishment of language, when treated as pure utility, that prevents the right ontological questions to be formulated. Heidegger's thought is used throughout this thesis; however, Groys' use here, to argue that words become free and equal, is misleading. Even if early search engines treated all words equally, contemporary incarnations have established a nuanced semantic structuring. The most recent iteration of this structuring is the implementation of Google's Knowledge Graph, launched in 2012, which draws from data correlations to build up a semantic map of words and phrases (see fig. 11).

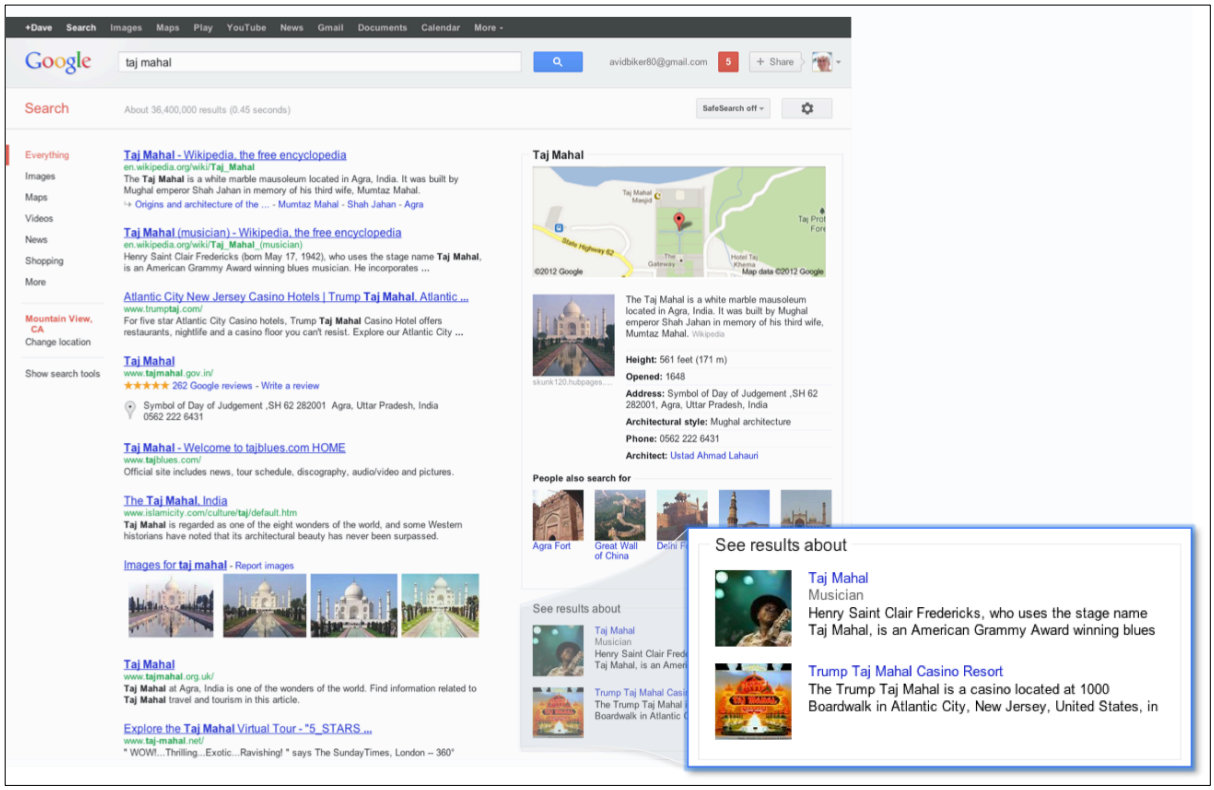


Figure 11. One of Amit Singhal's examples of the semantic search of Knowledge Graph, taken from his 2012 blog post "Introducing the Knowledge Graph: Things, Not Strings".

"Knowledge Vault" (Dong et al.) is an academic paper from a team at Google that explains part of the technical details, while Singhal's 2012 official Google blog post outlines the implications for users. Singhal describes how Google's Knowledge Graph categorises words and phrases as object or linked facts about objects so as to better categorise associations between query search terms. When Knowledge Graph launched, Singhal announced that it contained "500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects" ("Introducing knowledge Graph"). In 2016 Jeff Jarvis, author of *What Would Google Do?*, reported the number of facts held in Google's Knowledge Graph had increased to over 70 billion. Although Google's web index, from which it draws its search

results, is becoming increasingly structured many of Groys' conclusions still hold, in particular, the way in which queries are linguistically structured very differently from questions, the details of which constitute the following section. However, Groys' claim that the ties between words have been cut has never been true of search engines, even in their earliest days before the more recent advances in semantic search. Examples above, from the metaword tags used by AltaVista to the hyperlink measurements PageRank, represent active linguistic and extra-linguistic grammars. That HTML standards, as well as reading expectations, have directed the layout norms of webpages to reproduce various skeuomorphic resonances means that languages function in a similar way as they do offline. MacCormick's malaria example, demonstrating how AltaVista established a way of taking account of the distances between words, means that the kinds of pages that rank highly are those which the grammar and word order reflect the kinds of grammar and word order of queries. These relationships might not look like traditional grammar, but even in the early stages of search engines, structural rules that govern language operated in direct ways to enforce striations of meaning. In addition, as outlined above, the kinds of pages that PageRank designates as important are often those of established institutions; consequently, very traditional forms of grammar are prioritised and encourage many kinds of unestablished websites to follow these pre-digital norms. There are exceptions of course, but generally speaking grammar is still an important part of the search engines, the web, and the interaction between the two.

3.0 A Brief History of Search Engine Interface

The origins of modern search engines developed from search portals. Before the technological advances that enabled indexing and ranking to be performed algorithmically, traversing the web was accomplished through human curation of content and portals that offered a range of suggestions of websites to visit. In addition to the technical barriers to modern search engine interfaces, early sites operated at a time before the establishment of relatively standardised expectations regarding what the web was and how it could be used. A comparison between these established layouts and Google's sparse homepage reveals more than a difference in visual aesthetics; Google's homepage gave users no indication of what they might search for, or even how they might search. Ask Jeeves gave suggestions of what to search for, HotBot foregrounded a little insight into how a user's query was being interpreted, Excite offered categories much like a Dewey Decimal system, Google, however, gave users no such direction (see fig. 12-15).

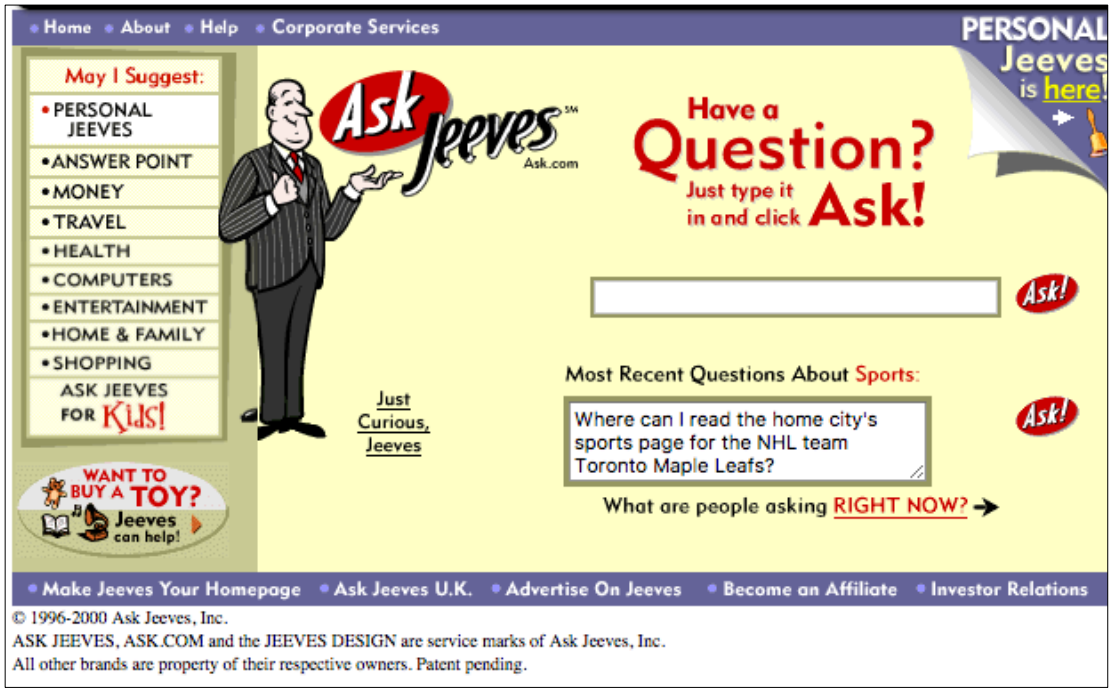


Figure 12. Ask Jeeves (ask.com), as of 29/02/2000, accessed via the Way Back Machine.

Figure 13. Hotbot, as of 10/12/1997, accessed via the Way Back Machine.

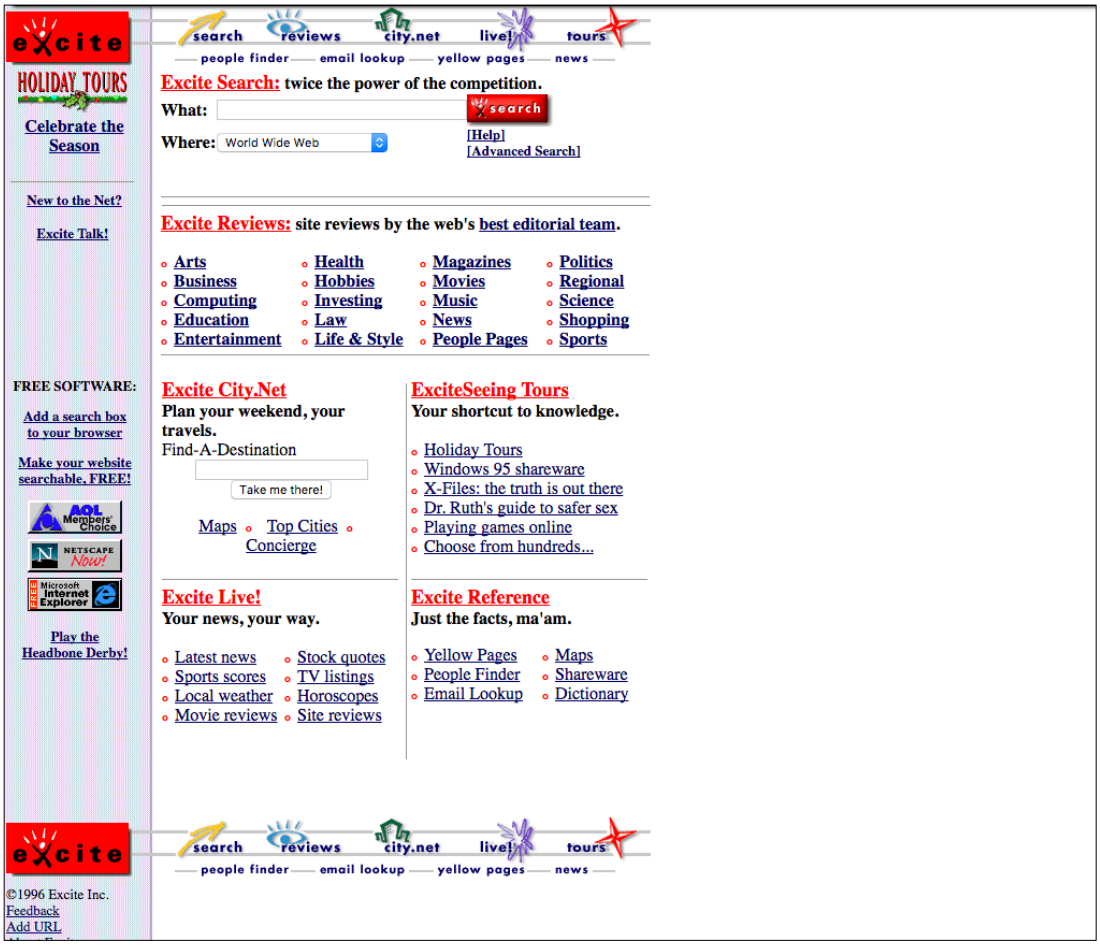


Figure 14. Excite, as of 19/12/1996, accessed via the Way Back Machine.

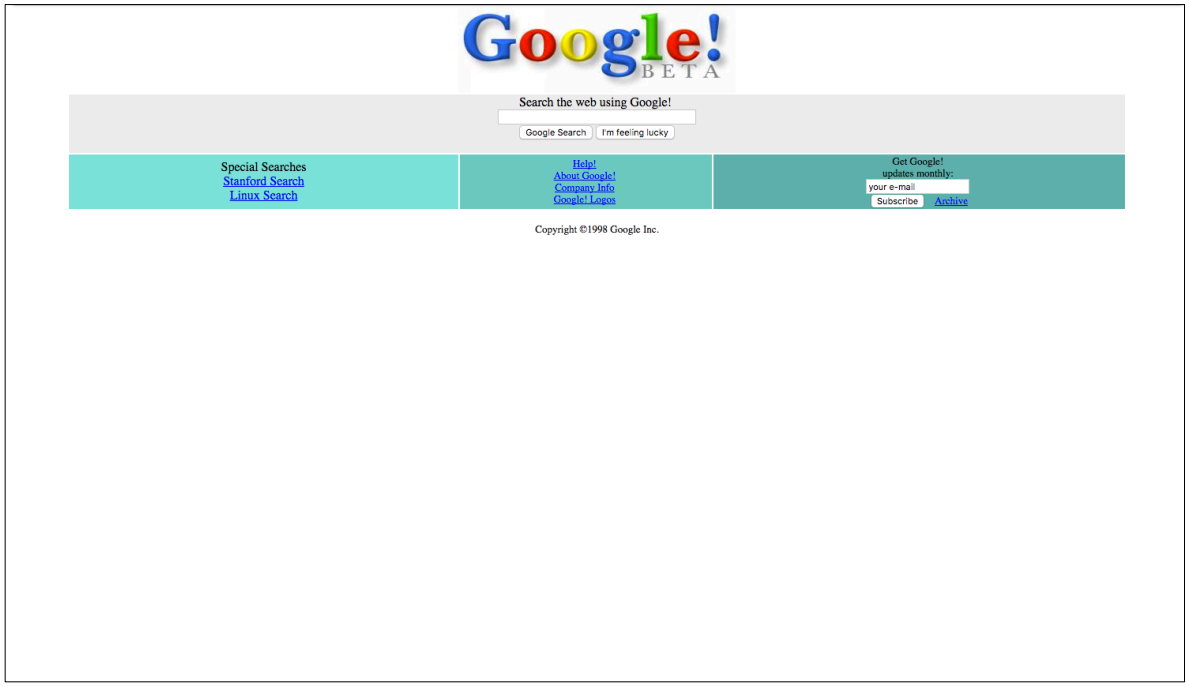


Figure 15. Google.com, as of 02/12/1998, accessed via the Way Back Machine.

This leads us to a key question of this chapter: how do people search Google? Even if there is a significant difference between the way that grammatically correct questions and search engine queries function, this does not necessarily tell us anything about what users actually type into Google's search box, or any other search engine for that matter. Google do not release search records, preferring instead to offer a range of tools listed under Google Trends (formally Google Zeitgeist), which provide some insight into the popularity of search topics, as well as currently searched queries, which represent a very small sample size. These features are discussed below. However, although Google have not released any search records, other search engines have. By surveying the literature of such studies, we can gain some level of insight into actual search queries and generalised search behaviours, as well as details such as average query lengths. Although there are a number of search engine query logs studies, their results and conclusions have not travelled outside of a niche IR sub-discipline into wider search engine studies. To my knowledge, the following section is the first time that such research has been used to inform search engine research outside of IR.

4.0 Transaction Log Analysis

Transaction log analysis (TLA) is a statistical method for researching corpuses comprised of some form of digital transaction, for example, queries used to search a particular database. A useful subcategory of TLA, for this study, is query log research. Google do not release transaction logs[12] but there has been research conducted on older non-Google logs, captured data between 1997 and 2006 (see Silvestri 17 for an overview), which provide an insight into query formation in this nine-year period. In addition, extrapolating from the findings provides a better prediction of the kinds of queries submitted to Google than simple guesswork based on anecdotal evidence of personal search strategies. There are, however, a number of limitations to the applicability of this related research; in particular, these search logs cover a time before Autosuggest or Autocomplete tools became a standard part of the query process in 2008. I argue, in chapter three, that the effect of Autocomplete and Autosuggest on query composition is significant and therefore the existing literature of TLA of search engine queries provides useful historical insight but might not accurately describe the kinds of queries used after 2008. An underlying question of this thesis concerns the agentic negotiations between human actors, wider social influence, and technologies that seek to direct and control in a cybernetic sense. Given this, speculation on how query composition has changed post-2008 draws on a much wider debate regarding the degree to which human-computer interactions (HCI) are shaped by technological affordances. Those that

[12] There are two reasons for this. First, query log data can impact the competitiveness of a particular search engine. Second, depending on the level of data released, users can often be identified through their searches, as was highlighted by AOL's 2006 log release, which resulted with the firing of multiple employees. It is therefore not a coincidence that 2006 is the final release of query logs by any search engine used in TLA research, for further details, see Silvestri "Mining Query Logs: Turning Search Usage Data into Knowledge" 11-15.

see HCI as dictated by unconstrained human freedom might conclude that contemporary search patterns would closely reflect those pre-2008. However, those who approach HCI with an attitude that the influence of design dictates much of a user's engagement might argue that contemporary patterns follow more directly from the form of the current search environment. Even if this second perspective is closer to the truth, technology requires a set of learned behaviours, embedded within social environments, which can become solidified over time and therefore the historical conventions of query input can have a strong influence of current search engine engagements.

Reviewing and summarising the TLA research on search engine queries allows us to draw a number of conclusions regarding the following topics: the types of goals or functions queries can be in service of; the average length of queries; the amount of success users have finding webpage links that satisfy their query; the number of results that are actually considered by users; certain correlations between time of day, length of query, and number of results considered; and the kinds of content users search for. This section outlines each of these and then draws on some insights into contemporary search queries as formulated for Google.

4.1 Goals, Functions and Intentions

People use search engines for a range of activities; sub-categorising these, either by academic researchers or programmers, limits the open-ended nature of web-based search. A key characteristic of search engines is that they can be used for a wide range of different tasks, an issue that will further be addressed in chapter two. Andrei

Broder argues in “A Taxonomy of Web Search” (2002) that due to its background in information retrieval, the academic understanding of web queries is biased towards defining them as “inherently predicated on users searching for information [even though] the need behind a web search is often not informational” (3). Following a query log analysis of queries submitted to AltaVista combined with a multiple-choice survey of AltaVista users,[13] Broder concluded that queries could be separated into three kinds: “1. Navigational: The immediate intent is to reach a particular site. 2. Informational: The intent is to acquire some information assumed to be present on one or more web pages. 3. Transactional: The intent is to perform some web-mediated activity” (5). The conclusion of the survey and log analysis provided the following breakdown of types of searches (see fig. 16).

Type of query	User Survey	Query Log Analysis
Navigational	24.5%	20%
Informational	?? (estimated 39%)	48%
Transactional	> 22% (estimated 36%)	30%

Figure 16. A table taken from Broder showing his query classifications and the percentage breakdown of the query functions (6).

This categorisation system has been adopted and adapted by subsequent researchers. For example, in 2004, Daniel Rose and Danny Levinson made the

[13] The work was carried out while Broder was working at AltaVista. It is common for this research to be conducted by employees of a particular search engine, as, even before AOL’s 2006 log release heightened concerns regarding privacy and the accompanying litigation, market competitiveness minimized publication of query logs.

following subcategorisations (see fig. 17) and found that their sample had much higher rates of “informational” queries and lower numbers of “navigational” and “transactional” than Broder (see fig. 18).

SEARCH GOAL	DESCRIPTION	EXAMPLES
1. Navigational	My goal is to go to specific known website that I already have in mind. The only reason I'm searching is that it's more convenient than typing the URL, or perhaps I don't know the URL.	aloha airlines duke university hospital kelly blue book
2. Informational	My goal is to learn something by reading or viewing web pages	
2.1 Directed	I want to learn something in particular about my topic	
2.1.1 Closed	I want to get an answer to a question that has a single, unambiguous answer.	what is a supercharger 2004 election dates
2.1.2 Open	I want to get an answer to an open-ended question, or one with unconstrained depth.	baseball death and injury why are metals shiny
2.2 Undirected	I want to learn anything/everything about my topic. A query for topic X might be interpreted as "tell me about X."	color blindness jfk jr
2.3 Advice	I want to get advice, ideas, suggestions, or instructions.	help quitting smoking walking with weights
2.4 Locate	My goal is to find out whether/where some real world service or product can be obtained	pella windows phone card
2.5 List	My goal is to get a list of plausible suggested web sites (I.e. the search result list itself), each of which might be candidates for helping me achieve some underlying, unspecified goal	travel amsterdam universities florida newspapers
3. Resource	My goal is to obtain a resource (not information) available on web pages	
3.1 Download	My goal is to download a resource that must be on my computer or other device to be useful	kazaa lite mame roms
3.2 Entertainment	My goal is to be entertained simply by viewing items available on the result page	xxx porno movie free live camera in l.a.
3.3 Interact	My goal is to interact with a resource using another program/service available on the web site I find	weather measure converter
3.4 Obtain	My goal is to obtain a resource that does not require a computer to use. I may print it out, but I can also just look at it on the screen. I'm not obtaining it to learn some information, but because I want to use the resource itself.	free jack o lantern patterns ellis island lesson plans house document no. 587

Figure 17. A table representing “The Search Goal Hierarchy” taken from Rose and Levinson’s “Understanding User Goals in Web Search”, which presents a subcategorisation of Broder’s original divisions (15).

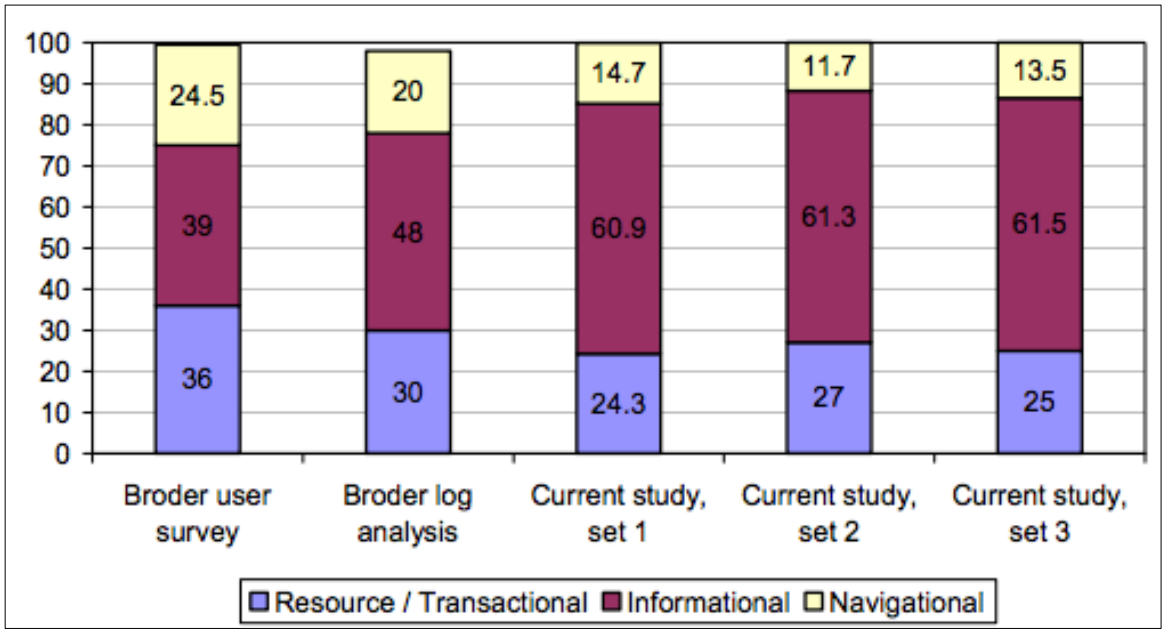


Figure 18. Graph taken from Rose and Levinson representing a "comparison of Broder's search taxonomy to our top-level goals. Resource and informational results in the first column are Broder's estimates. Results do not total 100% due to rounding error" (18).

Rose and Levinson stress the difficulties of concluding the function of different searches for a number of reasons. However, they stand by their claim, made in this paper as well as in other research,[14] that "goal-sensitivity will be one of the crucial factors in future search user interfaces" (13), meaning that search engine success is based on probabilistic assumptions of the goals of users. This kind of attitude draws us back to the infamous description of search engines as a "database of intentions" (Battelle 2) mentioned earlier, the outlook of which assumes that if search engines can correctly guess the goal of a user's query, the information required to produce their search results page flows unproblematically.

[14] See Daniel Rose "Reconciling Information-Seeking Behavior with Search User Interfaces for the Web".

To foreground their assertion that intention is the key factor for successful results, Rose and Levinson present an example of a search for a single word [ceramics], which as we will see shortly such short, open-ended queries are common:

> The “why” of user search behavior is actually essential to satisfying the user’s information need. After all, users don’t sit down at their computer and say to themselves, “I think I’ll do some searches.” Searching is merely a means to an end – a way to satisfy an underlying goal that the user is trying to achieve. (By “underlying goal,” we mean how the user might answer the question “why are you performing that search?”) That goal may be choosing a suitable wedding present for a friend, learning which local colleges offer adult education courses in pottery, seeing if a favorite author’s new book has been released, or any number of other possibilities. In fact, in some cases the same query might be used to convey different goals – for example, the query “ceramics” might have been used in any of the three situations above (assuming it is also the title of the book in question). (13)

Although Rose and Levinson stress the importance and ambiguity of search goals, I do not think they go far enough. Those in the discipline of English know how original intention can be a dangerous and unproductive red-herring and so can simplifying the actions of individuals down to clear, intentional and rational arrangements. Although query transaction logs *do* provide information as to whether or not a user followed a suggested result URL as well as what the page was, it is quite an assumption to make to argue that this provides an insight into the original intention

for searching that query. Perhaps a user makes a query intending it as an "informational" request (looking up the history of timepieces) but ends up following a "transactional" category link (clicking on an online watch shop). Part of a search engine's technological affordance is its open-ended nature; users do not necessarily need any kind of aim or functional forethought before typing a query. To take a metaphor from quantum mechanics, surveys such as Broder's when asking users their reason for making a query, might collapse a superposition of searching outlooks into an overly neat singular goal that does not fully reflect the overall process. Similar criticisms can be made of the seemingly more impersonal and objective data that link the query made to the result followed; when a user searches an open-ended query the result they follow says a lot more about the overall architecture of the web and the intrinsic biases of search engine ranking than it does their needs and desires as a user.

Jansen et al. use Broder's original three categories to structure their TLA findings,[15] suggesting that "more than 80% of Web queries are informational in nature, with about 10% each being navigational and transactional" ("Determining the Informational" 1251). They reflect on this seeming shift towards an increase in informational queries in each study since Broder's: rather than seeing it as an actual shift in search behaviour their conclusion is that "variation in reported percentage of navigational and transactional queries may be related to the size of the samples used in prior studies (which were much smaller than we used in this research)" (1262). Similar to the critique presented above concerning intention, I feel here that sample size may not be the only factor; navigational and transactional are much more clearly defined categories than informational and I think that there is a

[15] For examples, see Jansen et al. "Determining the Informational, Navigational, and Transactional Intent of Web Queries" (2008) and "Patterns of Query Reformulation During Web Searching" (2009).

tendency in large corpus studies for ambiguity to be collapsed down into the broadest category. They do provide a clear three-tier system of categorisation (see fig. 19) with some useful examples that help situate the reader to the scope of possible search queries, but I find it convinces me further that dividing queries between informational, navigational and transactional is not as useful a division as it might seem at first.

Definitions of classifications of Web queries

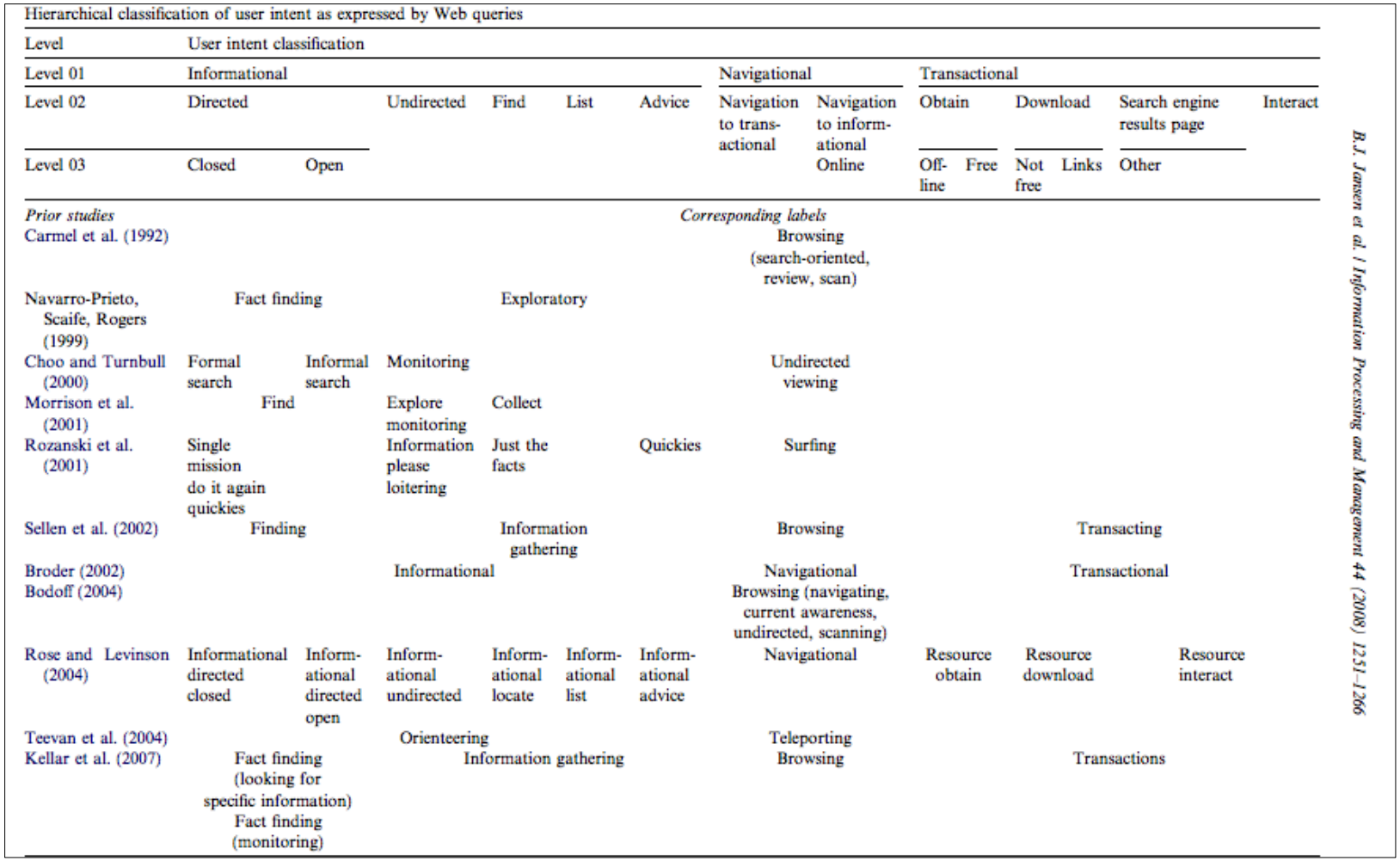
Levels	Examples of queries
Level one	
• (I) Informational: queries meant to obtain data or information in order to address an information need, desire, or curiosity • (N) Navigational: queries looking for a specific URL • (T) Transactional: queries looking for resources that require another step to be useful	• Child labor law • Capitalone • Buy table clocks
Level two	
• (I, D) Directed: specific question • (I, U) Undirected: tell me everything about a topic • (I, L) List: list of candidates • (I, F) Find: locate where some real world service or product can be obtained • (I, A) Advice: advice, ideas, suggestions, instructions • (N, T) Navigation to transactional: the URL the user wants is a transactional site • (N, I) Navigation to informational: the URL the user wants is an informational site • (T, O) Obtain: obtain a specific resource or object • (T, D) Download: find a file to download • (T, R) Results page: obtain a resource that one can printed, save, or read from the search engine results page • (T, I) Interact: interact with program/resource on another Website	• Registering domain name • Singers in the 1980s • Things to do in hollywood ca • PVC suit for overweight men • What to serve with roast pork tenderloin • match.com • yahoo.com • Music lyrics • mp3 downloads • (The user enters a query with the expectation that 'answer' will be on the search engine results page and not require browsing to another Website) • Buy table clock
Level three	
• (I,D, C) Closed: deals with one topic; question with one, unambiguous answer • (I,D, O) Open: deals with two or more topics • (T, O, O) Online: the resource will be obtained online • (T, O, F) Off-line: the resource will be obtained off-line and may require additional actions by the user • (T, D, F) Free: the downloadable file is free • (T, D, N) Not free: the downloadable file is not necessarily free • (T, R, L) Links: the resources appears in the title, summary, or URL of one or more of the results on the search engine results page • (T, R, O) Other: the resources does not appear one of the results but somewhere else on the search engine results page	• Nine supreme court justices • The excretory system of arachnids • Airline seat map • Full metal alchemist wallpapers • Free online games • Family guy episode download • (As an example, a user enters the title of a conference paper in order to locate the page numbers, which usually appear in one or more of the results) • (As an example, a user enters a query term to check for spelling with no interest in the results listing)

Figure 19. Table taken from Jansen et al. "Determining the Informational" demonstrating their three-tier classification system (1260).

Jansen et al. also suggest dividing search intent into "affective, cognitive, or situational goal[s]" (1255), a categorisation system that seems to reflect a fuller

sense of intent, although they do not provide a statistical model for separating the corpus into these categories and therefore do not elaborate on it further. Chapters two and three of this thesis open up some of the intersections between affective, situational, and cognitively extended recall in the context of search engines and other technologies. Jansen et al. provide a useful diagram of categories used in previous studies (see fig. 20).

In summary, the above categories of subdividing query types can be a useful insight into broad trends of how users search. This can usefully broaden the categories that we *as individuals* draw on when we try to define what search engines are for. We may conclude that between 2002 and 2009 the research community came to a consensus that informational queries outnumbered other types of queries to a greater extent than was originally proposed in Broder in 2002. However, as stated above, what such a claim really tells us about how people formulate queries and more generally what role users conceive search engines to be fulfilling, might be limited.



Hierarchical classification of user intent as expressed by Web queries

Level	User intent classification														
Level 01	Informational						Navigational		Transactional						
Level 02	Directed		Undirected	Find	List	Advice	Navigation to trans-actional	Navigation to inform-ational	Obtain		Download		Search engine results page		Interact
Level 03	Closed	Open						Online	Off-line	Free	Not free	Links	Other		
Prior studies							*Corresponding labels*								
Carmel et al. (1992)								Browsing (search-oriented, review, scan)							
Navarro-Prieto, Scaife, Rogers (1999)	Fact finding			Exploratory											
Choo and Turnbull (2000)	Formal search	Informal search	Monitoring					Undirected viewing							
Morrison et al. (2001)	Find		Explore monitoring	Collect											
Rozanski et al. (2001)	Single mission do it again quickies		Information please loitering	Just the facts		Quickies		Surfing							
Sellen et al. (2002)	Finding			Information gathering				Browsing				Transacting			
Broder (2002)			Informational					Navigational				Transactional			
Bodoff (2004)								Browsing (navigating, current awareness, undirected, scanning)							
Rose and Levinson (2004)	Informational directed closed	Inform-ational directed open	Inform-ational undirected	Inform-ational locate	Inform-ational list	Inform-ational advice		Navigational	Resource obtain		Resource download		Resource interact		
Teevan et al. (2004)			Orienteering					Teleporting							
Kellar et al. (2007)	Fact finding (looking for specific information)		Information gathering					Browsing				Transactions			
	Fact finding (monitoring)														

B.J. Jansen et al. / Information Processing and Management 44 (2008) 1251–1266

Figure 20. Table taken from Jansen et al. « Determining the Informational» , which surveys a range subdivisions, drawn from different studies, regarding query types (1259).

4.2 Length of Queries

In general, the lengths of queries from all of the TLA samples are short. This reinforces the distinction established earlier that queries are of a very different kind of linguistic form to questions. Even if the reasoning for this is not because users understand how search engines function, queries are usually too short to be phrased as questions. See fig. 21 for a table of the average query lengths from different TLA studies.

Study	Date of query log collection	Average number of terms per query
Spink et al. (2002)	Three combined logs: 1997, 1999, 2001 – Excite Logs	2.4
Silvestri (2010)	Two combined logs 1998 – Excite Log 1998 – “private” AltaVista Log	2.35
Özmutlu et al. (2004)	1999 – Excite Log	2.9
Silvestri (2010)	2001 – “public” AltaVista Log	2.55
Jansen et al. (2009)	2005 – Dogpile Log	2.79
Zhang et al. (2009)	2006 – Dogpile Log	2.9
Pass et al. (2006)	2006 – AOL Log	3.5

Figure 21. Table of average terms per query, compiled by the author from various studies. Listed chronologically in order of the “date of query log collection”.

The mean averages of the query lengths are very short; most of the averages surveyed were between two and three terms per query. This suggests that users are not conforming to the traditional question grammar and are instead using keywords

arranged ungrammatically. The 2006 AOL log represents the highest average at 3.5. However, because it represents the last log to be released and studied, whether it is part of a general trend towards longer queries is not known. None of the above studies cover the relationship between queries and grammatically formed questions. A cursory look through any of the publicly available logs does reveal that many users still frame their queries in the form of questions. Fig. 22 shows a small sample.

how to choose cookware
how to make money from home
why is the poor so happy
why is city ripping up 40 million dollar trolley line
what is echocardiogram
what a spa administration is like
who sings jet city woman
who discovered the gene for huntington's disease
where are the basal ganglia
where to buy earthworms
when the civil war ended
when can you tell a fish is going to have baby
how to texture drywall
how to know if your child is on drugs

Figure 22. A randomly selected sample of queries framed as questions, taken from the 2006 AOL Query Log release. The log is currently available through McGill's Centre for Intelligent Machines website, see the bibliographic entry "AOL 2006 Query Log" for the URL of the full log and Pass et al. for its analysis.

Given the average query lengths reported in fig. 21, we can conclude that queries framed as questions are unusual due to their length. Looking at fig. 22 also reinforces the point made earlier about how search queries are open-ended and that matching the query to the result followed by the user does not necessarily inform us about intention. Take the query [where to buy earthworms], framed in this way we can describe this query as transactional. However, it would be more likely that a user would enter a shorter and more open-ended query, such as [earthworms] (figs. 23 and 24 compare the results from Google). The shorter query of [earthworms], when searched today, provides many more informational results; these results may change the overall nature of the query. Perhaps the user searching for [earthworms] has an initially transactional intention, wanting to buy earthworms to feed to a pet. However, when the results for such a query are mainly informational nature, it is likely they might follow an informational result, learn more about earthworms and as a result choose not to buy earthworms after all. In doing so, an informational result has fulfilled a transactional enquiry. There are many other ways in which even queries with the clear, definable intentions can shift and change. There is no clear way to measure the changing nature of queries and therefore represents a challenge to TLA research that foregrounds intention as a metric. The length of queries and their relationship to traditional questions will be developed in more depth in chapter three.

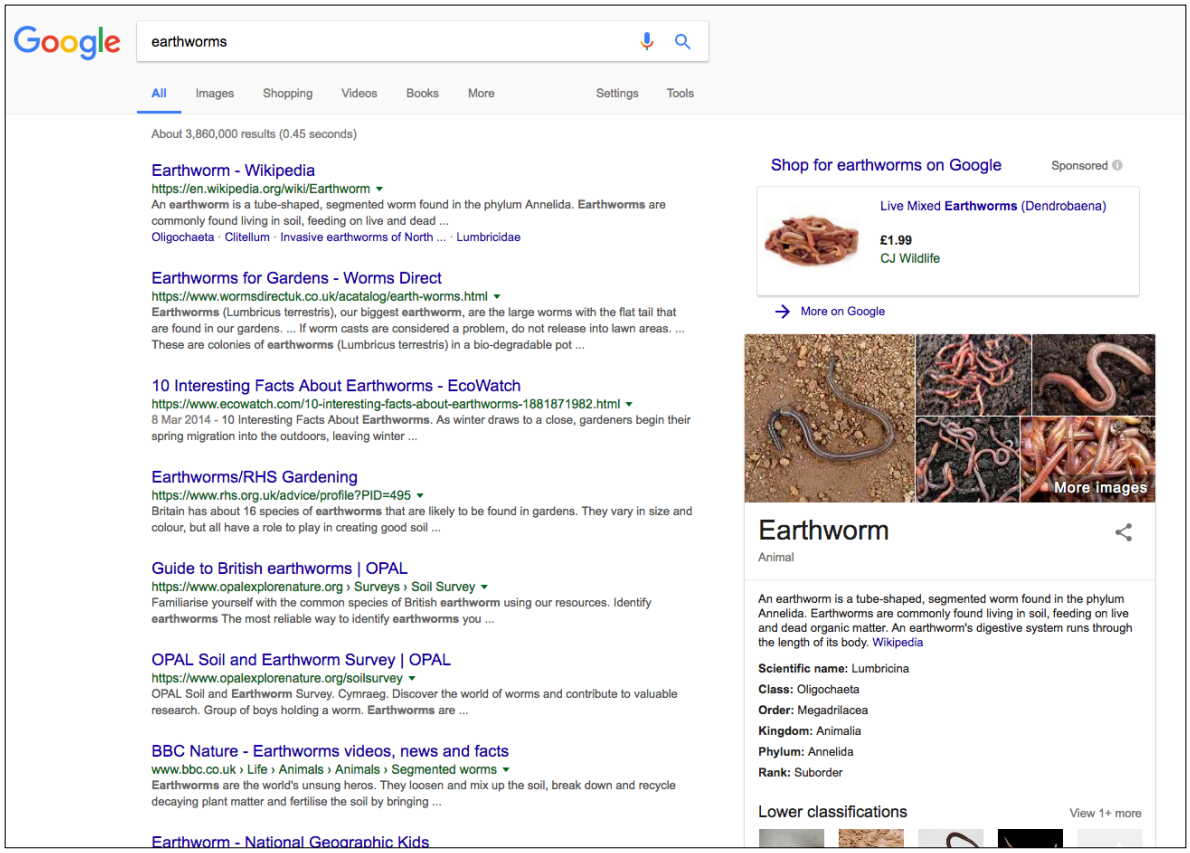


Figure 23. A Google search for [earthworms] performed 17/7/17.

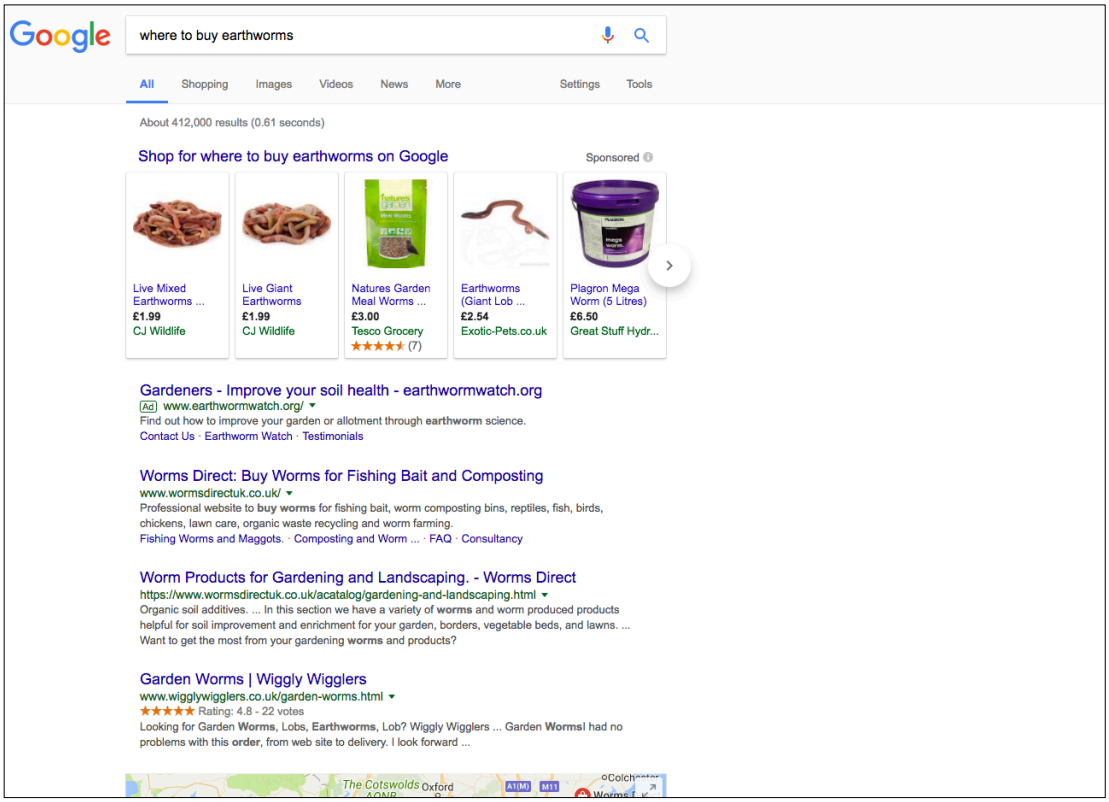
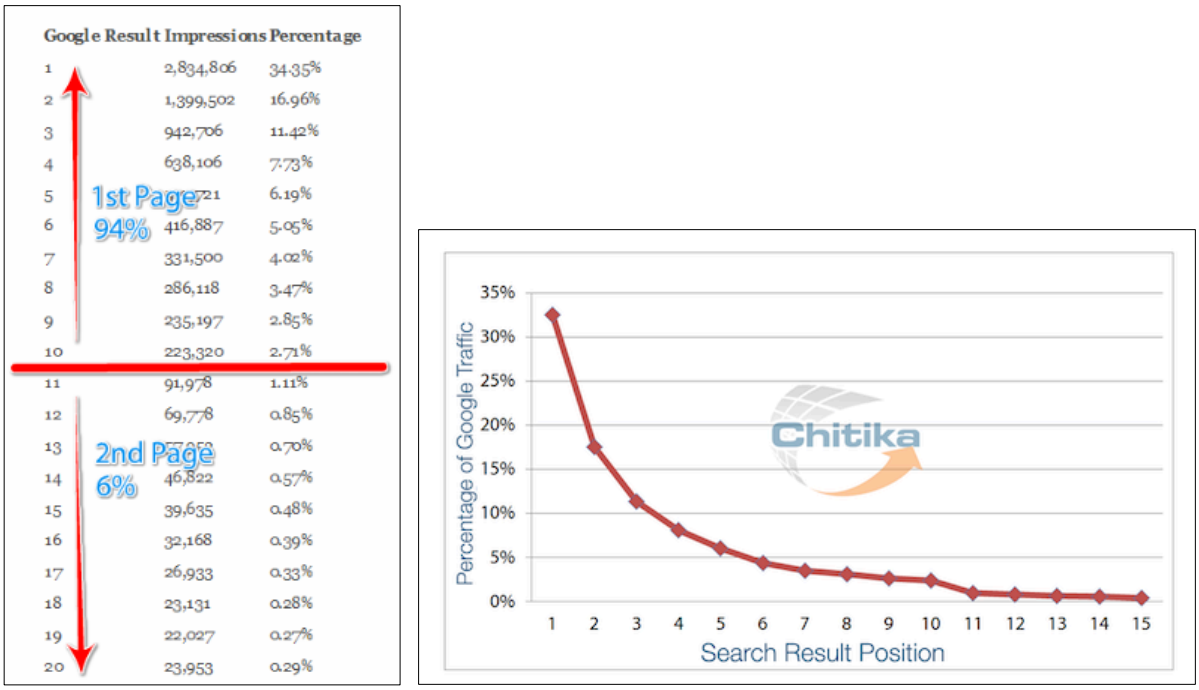


Figure 24. A Google search for [where to buy earthworms] performed 17/7/17.

4.3 Search "Success" and How Deep Searchers Dig

A review of the TLA literature provides us with an insight into how users deal with a search results page. A general noted trend is that users rarely follow results further than the first couple of results and almost never navigate past the first page of results. See, for example, figs. 25 and 26 for a 2013 study regarding the Google results that users follow; it demonstrates that more than 50% of the time the users studied followed either the first or second result and that page two results are relatively negligible.



Figures 25 and 26. Showing that users rarely navigate to Google's second page of results. Taken from Chitika Insights "The Value of Google Result Positioning".

This is not to say, however, that users' queries are always satisfied by these top results. Zhang et al. found that "approximately 40% of the time, searchers entered queries but did not click on any links in the SERP [search engine results page]" (243). This correlates with the findings of Jansen at al. in their study "Patterns of

Query Reformulation During Web Searching", in which they found that "Nearly 40% of query submissions were some sort of query reformulation" (1366). Fig. 27 shows the categorised reasons, for why queries had been reformulated.

Search patterns	Occurrence	%	Occurrence (excluding *New*)	% (excluding *New*)
New	964,780	63.34	–	–
Reformulation	126,901	8.33	126,901	22.73
Assistance	124,195	8.15	124,195	22.25
Specialization	90,893	5.97	90,893	16.28
Content change	65,949	4.33	65,949	11.81
Specialization w/reformulation	55,531	3.65	55,531	9.95
Generalization w/reformulation	54,637	3.59	54,637	9.78
Generalization	40,186	2.64	40,186	7.20
	1,523,072	100.00	558,292	100.00

Figure 27. Table taken from Jansen et al. "Patterns of Query Reformulation During Web Searching" (1366), showing a categorisation of the reasons why queries were reformulated.

An important aspect to note is that the 40% of reformulated queries does not include queries reformulated due to spelling errors. We can conclude that, generally speaking, if a user does not see a link that seems to satisfy their query in the first few results they are much more likely to reformulate their query than to keep looking through the other results, by scrolling down the page or moving to the second results page. These figures reinforce the earlier assertion that users have specific goal-oriented intentions; such a high number of query reformulations suggests that open-ended browsing may be uncommon. It also recontextualises the behaviour, outlined above, of users following one of the first results; because query reformulation is common, it may be that rather than mindlessly following the top link due to a lack of thought, users are actively engaged when judging the appropriateness of results. Reformulating a query requires greater cognitive attention than simply scrolling down a page so such a statistic may go against the dominant stereotype that people only follow the top result due to laziness or stupidity.

A related set of research that reinforces the conclusions drawn from these query TLA studies is that of eye-tracking research. This research uses eye-tracking technology to map where users look on a search result page (see fig. 28-31). These were conducted more recently and also draw from a much smaller sample size than TLA studies but they provide a very different way of measuring a related aspect of searching.

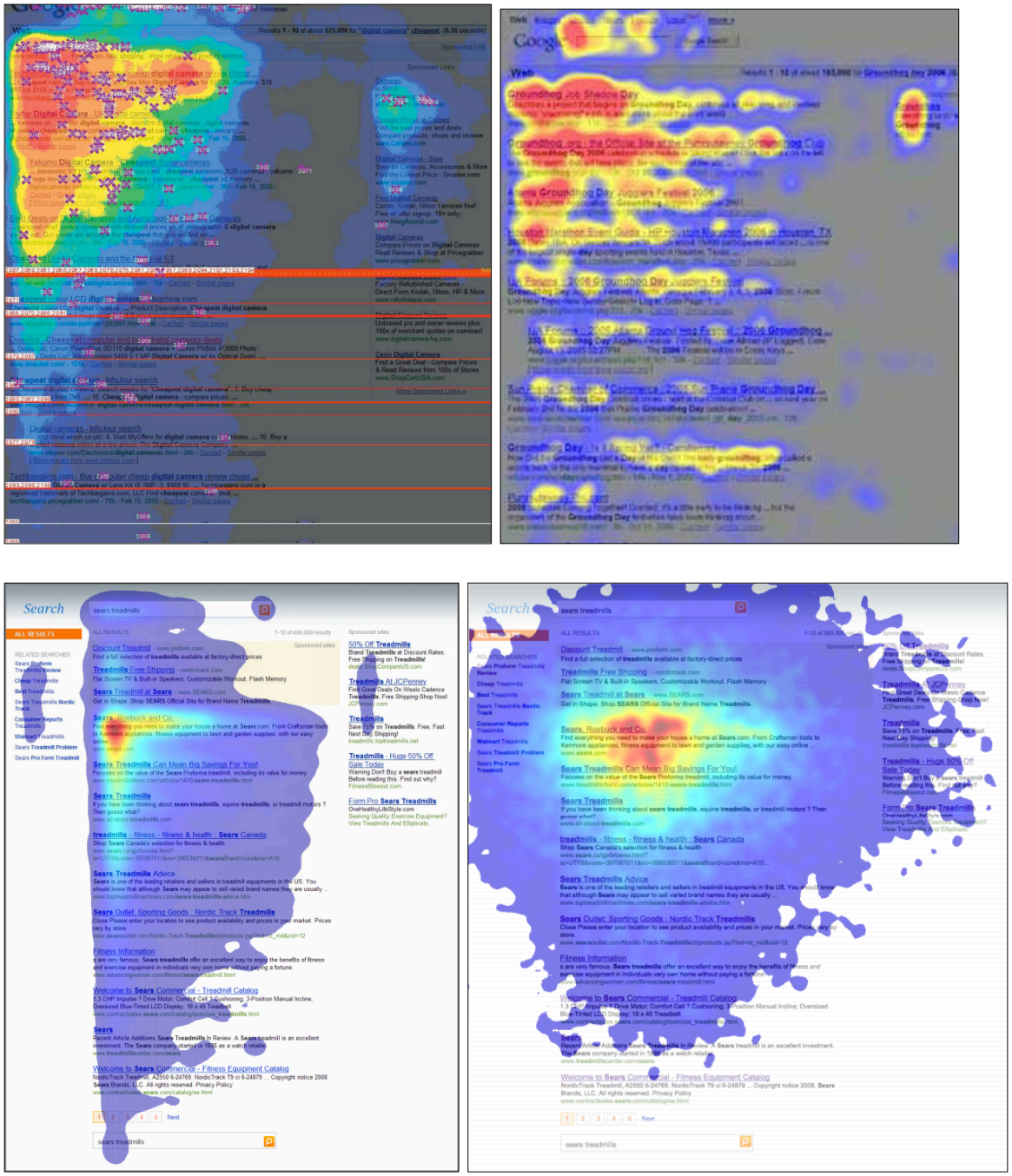


Figure 28-31. Taken from various tracking study by in which the eye movements of participants were recorded as they engaged with a Google results page. This image shows a heatmap of the aggregated data regarding where participants were looking. (From left to right) Hotchkiss et al. (7), Nielsen, Buscher (42), Papoutsaki et al. (7).[16]

[16] Figure 31 shows a different result to the earlier studies. The authors of the study propose that the wider field of visual attention, as opposed to the golden triangle or f-shape, is due to Google's contemporary way of presenting results that changes the layout based around the knowledge graph information.

4.4 Correlations

Taking into account all the TLA studies, most do not mention correlations. One correlation that is mentioned is in Zhang et al., who found that

> if people in a given period typed in fewer query terms, then they were more likely to click on the top ranked results in the following period. This could be one indication of a particular searcher need, most probably navigational searcher intent. Jansen et al. (2008) found that searchers with navigational queries clicked on higher ranked results than did searchers with informational and transactional needs. (244)

Correlations such as these are difficult to interpret. Zhang et al., assume that few search terms equates to navigational searching, building from the conclusions that Jansen et al. made in 2008, published in “Determining the Navigational”. However, another way to interpret this correlation is to argue that users without technical proficiency or much experience using search engines might both search using fewer queries and follow the top links, assuming them to be correct. User expertise is a crucial issue when exploring a technology and a lack of understanding can have a dramatic effect on how particular technologies, search engines for example, come to be used. Particular attention should be paid towards the ways in which children use search engines as this provides an insight into possible future usage. A further interpretation may be that searchers *do* browse without any intended goal, searching one term and simply visiting the highest ranked result. Without more evidence, making such claims relies on much guesswork and, as such, is a reminder of the

methodological difficulties of studying dynamic systems, especially those relying on proprietary technology and operated by companies that rarely release data. It is also misleading to treat such correlations as ones that would be easily understood with enough data; it may well be the case that Google engineers do not have a full understanding of such search behaviour. The longer a query is, the more evidence it provides for analysis, by algorithms and researchers. Many of Google's adaptations are ways of drawing more information out of a searcher, or their context; a clear example of this is Autocomplete, addressed in chapter three, which can be contextualised as a way for Google to nudge users away from short or one-word queries. However, due to the lack of query logs released after 2006, we can only speculate on the changing length of queries since that date.

4.5 Content of Search Queries

Finally, the TLA literature provides a broad map of the changing nature of the content of user's search queries. In "How Are We Searching the World Wide Web? a Comparison of Nine Search Engine Transaction Logs", Jansen and Spink provide a five-year overview (see fig. 32).

	Categories	1997 Excite (2414 queries) (%)	1999 Excite (2539 queries) (%)	2001 Excite (2453 queries) (%)	2002 AltaVista (2603 queries) (%)
1	People, places, or things	6.7	20.3	19.7	**49.3**
2	Commerce, travel, employment, or economy	13.3	**24.5**	**24.7**	12.5
3	Computers or Internet	12.5	10.9	9.7	12.4
4	Health or sciences	9.5	7.8	7.5	7.5
5	Education or humanities	5.6	5.3	4.6	5.0
6	Entertainment or recreation	**19.9**	7.5	6.7	4.6
7	Sex and pornography	16.8	7.5	8.6	3.3
8	Society, culture, ethnicity, or religion	5.7	4.2	3.9	3.1
9	Government	3.4	1.6	2.0	1.6
10	Performing or fine arts	5.4	1.1	1.2	0.7
11	Non-English or unknown	4.1	9.3	11.4	0.0
		102.9	100.0	100.0	100.0

Figure 32: Distribution of Excite and AltaVista general topic categories, taken from Jansen and Spink. Original note: “Bolded percentages indicate the highest ranked topic in a given year.”

Jansen and Spink document that there was “a steady rise in searching for People, Place or Things and Commerce, Travel, Employment or Economy, with decreased searching for Sex and Pornography and Entertainment or Recreation” (259). The decrease in “sex and pornography” related queries was also echoed in other studies; however, it draws up a more general difficulty in how changes in query topics can be interpreted. Jansen and Spink conclude that the “decrease in sexual searching as a percentage of overall Web searching” is part of an overall trend:

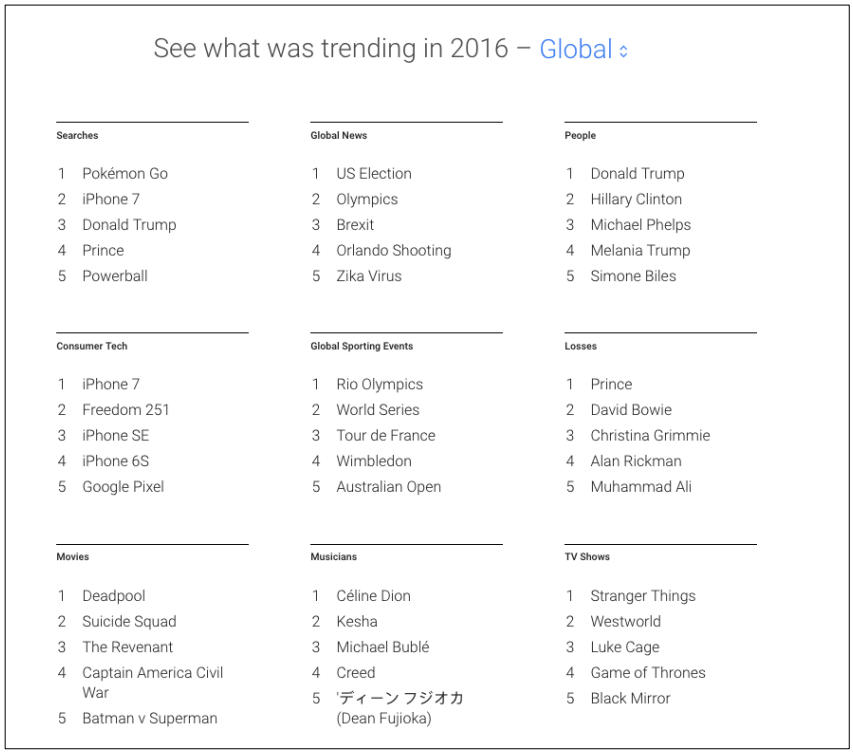
> towards using the Web as a tool for information or commerce, rather than entertainment. […] This analysis certainly confirms survey and other data that the Web is now a major source of information for most people (Cole et al., 2003; Fox, 2002). There is increased use of the Web as an economic resource and tool (Lawrence & Giles, 1999; Spink et al., 2002), and people use the Web for an increasingly [sic] variety of information tasks (Fox, 2002; National Telecommunications & Information Administration, 2002). (260)

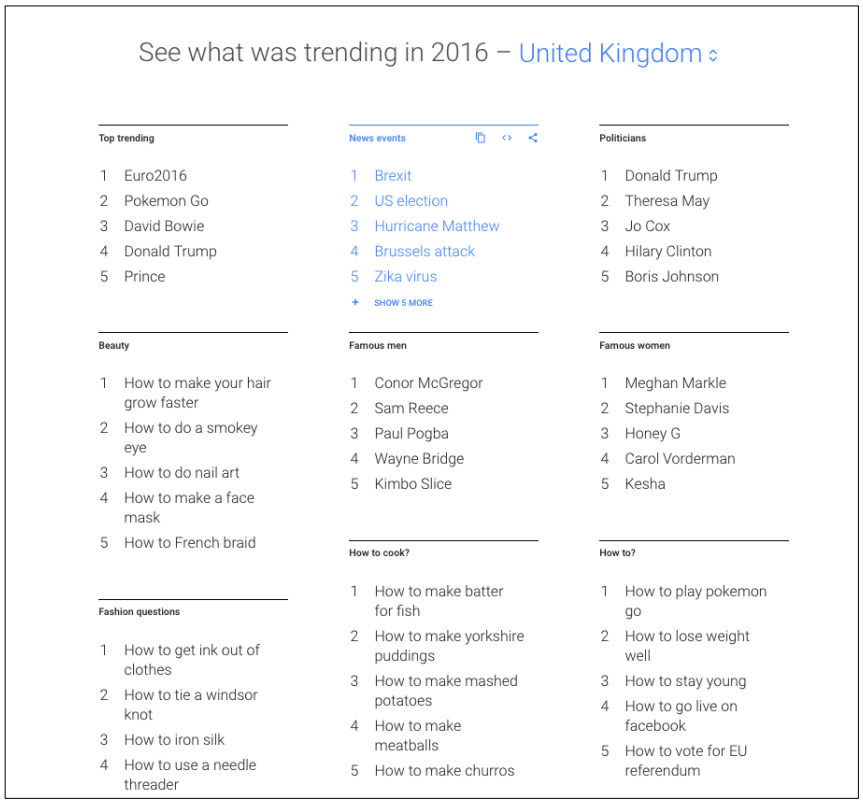
Jansen and Spink only outline the percentage changes of topics for search queries. Therefore, we cannot know whether the decrease in the dominance of sexual searches is a decline in the number of searches or whether the number of these searches has grown but on a much smaller scale than “people, places, or things”, a rather broad category. This result might actually point to the growing sophistication of users’ knowledge of the web and sites that are available for such content. It could also measure a wider trend towards a public awareness that queries are logged and therefore a moderation in their online behaviour.

This brings us to a wider point, even when TLA seems to provide a lot of information much of the nuance of searching a specific query is lost. It is unfortunate that the only query logs that researchers have had access to are between 1997 and 2006. However, in many ways, TLA oversimplifies many of the complexities of search queries and the data only allow for certain kinds of research questions to be posed. The conclusions above can serve as a useful guide that can underlie investigation of search behaviour post-2006. Nevertheless, that studies cannot use TLA to study search behaviour from 2006 to the present frees up, to some extent, the kinds of research questions that can be raised.

The most comparable resource available for Google search queries is Google Trends (previously known as Google Zeitgeist), which is a service that provides an “unbiased sample of […] Google search data” (Rogers). In particular, Google Trends provides a graphical representation of particular search phenomena, such as popular search topics within a time frame. However, the academic value of this service is limited as the “data” that Google release are based on relative comparisons; all that can ever really be inferred are differences in volume of particular search topics, because the data relating to actual search figures are not released (see figs. 33 and

34). Trending is a measurement of how much a topic is being used in search queries compared to other times. Therefore, it is unlikely that “Stranger Things” (see fig. 33) was the most searched for television program, only that the relative difference between that term’s use in 2016 is much higher than in 2015 (given that the programme only first aired in 2016 this relative comparison tells us little).





Figures 33 and 34. Google Trends representing “what was trending in 2016 (global)” and “what was trending in 2016 (United Kingdom)”, respectively. See Google “Google’s Year in Search 2016”.

At best Google Trends is little more than gimmicky self-promotion and at worst has caused some serious misunderstanding when used as evidence in journalistic claims. An example of this followed the results of the 2016 UK European Union membership referendum, in which a number of @GoogleTrends tweets (see figs. 35 and 36 for two examples) were used by a range of journalists as a form of evidence.

Figure 35. @GoogleTrends tweet regarding “+250% spike in ‘what happens if we leave the EU’ in the past hour” after the Brexit Referendum results were announced 23/06/2016.

Figure 36. @GoogleTrends tweet listing the “top questions on the European Union in the UK since Brexit result officially announced” 24/6/2016.

The tweets highlighted the rise of particular Google search queries following the results of the referendum. These tweets were used by various journalists as “evidence” that a large number of the British population were uninformed about the EU when voting in the referendum. Example articles include the *Washington Post’s* “The British are frantically Googling what the E.U. is, hours after voting to leave it” (Fung), *The Independent’s* “What happens if we leave the EU? Google searches surge as people realise they don't know what Brexit actually means” (Griffin), the *Mail Online’s* “Google search spike suggests many people don't know why they voted for Brexit” (O’Hare) as well as similar articles by *USA Today* (Blumenthal), the *BBC* (Baraniuk), *Fortune* (Roberts), *Time* (Chan), *NPR* (Selyukh), *Ars Technica*

(Walton), *The Huffington Post* (Beres), the *Mirror* (Wilson), the *Metro* (Willis), *The Irish Times* (McMahon) and a number of other outlets. However, because Google Trends only provides the relative differences of search terms it is unclear how many individuals turned to Google for information after they had voted. Danny Page's "Stop Using Google Trends", published in direct response to the widely circulated claims of the above articles, outlined the lack of context and transparent data available for such claims to be made. Page's article includes a Google Trends graph (see fig. 37) created by Twitter user @sammich_BLT that demonstrates how even a relative increase of 250% may be relatively small compared to other searches and may not reflect a large number of individuals, as was reported by numerous news outlets. This example demonstrates the lack of transparency regarding the content of search queries submitted to Google. It also exhibits the way in which many news organisations actively diminish the public's information-literacy in their pursuit of maximising page views, a situation in which, as will be outlined in chapter five, Google is partly complicit.

Figure 37. A Google Trends graph comparing the relative search query volume of three search queries [What is the EU?], [Game of Thrones], and [Euro 2016] after the announcement of the UK EU referendum results. Image by @sammich_BLT on Twitter 24/06/2016 and cited in Danny Page's "Stop Using Google Trends".

Another false flag for understanding the content of search queries submitted to Google is Google's "Year in Search" videos (previously Google Zeitgeist and Year in review from 2010 onwards). These videos are released in coordination with Google Trends and represent Google's overall marketing strategy of placing their technology at the centre of world events, rather than actually presenting an insight into how users search Google. Two stills from "Google – Year In Search 2016", (fig. 38 and 39) represent this trend and highlight two much-searched events. In the context of these two events, releasing detailed information regarding how and what users search would prove very valuable to the academic community but may not echo the branding sentiments of Google.

Figures 38 and 39. These stills from Google's end of year video "Google – Year in Search 2016" show a search bar superimposed over footage of various world events.

5.0 Asking Questions About the World

The bulk of this chapter aimed to highlight the technical and linguistic differences between questions and search engine queries. Words function in different ways and are bound by a different kind of grammar when posed to a search engine. However, if the previous section's aim was to draw questions and queries apart from one another, this second section aims to outline their similarities. Questioning as a route to understanding has a long history and Google is very much a part of that tradition. Chapter two addresses this history primarily in terms of memory aids and technological structures of thought; this second section, therefore, aims to act as a platform to bridge the technical with the philosophical.

The questions posed in Plato's Socratic dialogues still resonate with the contemporary way in which we query search engines and usefully present challenging questions of that modern process. I will briefly introduce some of Plato's enquiry while also drawing from the philosophy of Hans-Georg Gadamer, both of whom continue to be put to work in the following chapters.

5.1 Meno's Paradox

In Plato's "Meno" dialogue Meno and Socrates discuss the nature of virtue. The dialogue begins with Meno asking whether virtue can be taught which then leads Plato's Socrates to begin his characteristic mental midwifery. Socrates is less interested in whether virtue can be taught, but rather, how Meno can define virtue in the first place. Meno provides a list of examples that demonstrate virtue, which Socrates is unsurprisingly unhappy with. What does virtue *really* constitute, Socrates

wishes to know, to which Meno replies: “How will you look for it, Socrates, when you do not know at all what it is? How will you aim to search for something you do not know at all? If you should meet with it, how will you know that this is the thing that you did not know?” (“Meno” 80d). Learning, for Meno, presents us with a paradox: either we know things or not. How would we begin our self-education from nothing? We do not know where to start and more importantly, we would not recognise something unknown to us, like virtue, even if we did find it. The paradox that Socrates outlines is placed here in a very different context but establishes some common ground between traditional questions and search engine queries. In fact, due to the way in which search queries are interpreted, as outlined above by searching an index for the words submitted, queries are trapped in Meno’s paradox at a deep level. Questions function by prompting some aspect of, as yet, unknown dialogue. Search queries, although they might ask any manner of things from the sacred to the profane, lead users into a trap built from their own language. The topics of enquiry and the phrasing of ideas can only build outward from a searcher’s own linguistic framework. A key argument of this thesis is that a user’s search capacity is a representation of their situatedness in language and culture. This does not represent Socrates’ solution to Plato’s dialogue; his answer is an appeal to the immortality of the soul, which lays the foundation for Plato’s theory of anamnesis. Anamnesis will be discussed in the following chapter. The conclusion that I want to take from “Meno”, at this stage, is that enquiry, through the form of questions, is an attempt to open ourselves up to alterity.

5.1.1 Gadamer's Hermeneutics

A useful model for a subject opening up to alterity is Gadamer's sense of an historically situated hermeneutics; in reference to R. G. Collingwood, Gadamer argues: "We can understand a text only when we have understood the question to which it is an answer" (379). Hermeneutical understanding, then, is a method of formulating the right questions: questions that are inherently part of the text in question. For Gadamer this attempt for "understanding is always the fusion of … horizons" (317) by which he means that all understanding emerges in the light of pre-existing prejudices: our situatedness. These horizons are the relationship an interpreter takes to a question, an object, or a text. To properly ask a question is to anticipate and situate one's perspective as in line as possible with the area of enquiry. Described in this way, a search engine query is even more an attempt to align with a horizon. Not only do queries have to foreground unfamiliar ideas, but they must do so in the linguistic register that a searcher believes the answer will be written. Search engines act as intermediaries, but rather than encouraging a dialectical relationship they allow users to view many different horizons. These horizons are, of course, ranked but their potential to foreground numerous contrasting or even contradictory perspectives and bodies of evidence is different to engaging with a single interlocutor. Search engines are always about a multiplicity of answers; structurally, each search result shows how enquiry is complex through the sheer number of results and practically infinite pages represented.

Gadamer's sense of understanding, which he borrows from Heidegger, is that it is a form of practical know-how. The effective use of a search engine represents practical know-how to a much greater degree than any of the epistemological content a user aims at. Here we are drawn back to Meno's paradox and to Plato's

distrust of the sophists and their routes to false knowledge. The successful approach to formulating a search engine query is by imitating the dialect of the source we want our answer to come from. As Jean Grondin argues:

> For Gadamer, Plato's clear intention was to show that we cannot attain the truth of things by words. By that, he sought to distance himself from the Sophists, who taught that we can be assured of domination over things by mastery of words (or rhetoric). For Plato, true knowledge should, on the contrary, seek to liberate itself from the empire of words by being directed to things themselves, that is to say, to the Forms. In saying that, Plato did not necessarily wish to deny that true philosophical thought continued to be deployed in language, but his essential point was that access to the truth is not given by words themselves, not their mastery. (131-2)

Plato's concern is that a mastery of language might shut down a useful dialectical enquiry. For Gadamer, because every act of interpretation involves the coming together of different horizons, an attempt to dominate this exchange will always end as a failure in understanding. This apprehension of illusory knowledge is mirrored in Plato's concerns with literacy, in that the failure to be open to questions presents a lack of knowledge. The echoes of Plato's distrust of technology, in this case writing, recur in various attitudes towards the impact of digital literacy and will be covered in chapter two. For the present context, Plato's notion that writing is a poor substitute for dialogue because it cannot answer back is key to understanding how queries, submitted to a machine, are different from questions, directed at an individual. It also speaks directly to the earlier TLA statistics that 40% of the time, users entered a

query but did not follow any of the links provided (Zhang et al.) and also that 40% of queries are reformulations of unsuccessful queries (Jansen et al. “Patterns of Query Reformulation”). If a user is faced with a page of results that do not reflect how they conceived their query, then that text, in a hermeneutical sense, asks a question of that user. This question relies on the user to provide a different query or apply an alternative understanding; search engine results do not help with this process in the way that an individual might. In reply to Phaedrus, Socrates outlines his concerns with literacy in the following way:

> writing shares a strange feature with painting. The offsprings of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You’d think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever. When it has once been written down, every discourse roams about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn’t know to whom it should speak and to whom it should not. And when it is faulted and attacked unfairly, it always needs its father’s support; alone, it can neither defend itself nor come to its own support. (Phaedrus 275 d-e)

The written word cannot talk back. The question is, is this true of search engines? In the most explicit and short-term ways they do not. This is why cultivating a deep understanding of how search engines work is key to an informed public and an

essential part of digital literacy. However, on a different scale, algorithmic results such as Google's are not as passive as Socrates describes the written word to be.

5.2 Algorithms in the Context of Questions and Queries

In the context of Google Search, each search feeds information back into Google's algorithms at a number of levels. This feedback of data changes the outcome of future searches as the system has new information for its algorithms to draw from due to the previous investigations. Every query entered that engages Google ranking system changes that system. The responsiveness of Google Search is due to the constant tests being carried out on its users. As Patrick Riley, a Search Quality Manager at Google, outlines in Steven Levy's *In the Plex: How Google Thinks, Works and Shapes Our Lives*:

> The mainstay of this system [of evaluation] was the 'A/B test,' where a fraction of users – typically 1 per cent – would be exposed to the suggested change. The results and the subsequent behaviour of those users would be compared with those of the general population. Google gauged every alteration to its products that way, from the hue of its interface colors to the number of search results delivered on a page. There were so many changes to measure that Google discarded the traditional scientific nostrum that only one experiment should be conducted at a time, with all variables except the one tested being exactly the same in the control group and the experimental group. 'We want to run so many experiments, we can't afford to put you in any one group, or we'd run out of people,' says a search quality manager [Patrick Riley]. On most

> Google queries, you're actually in multiple control or experimental groups simultaneously. Essentially *all* the queries are involved in some test. (61)

As well as each search being part of one or several experiments, every search provides data which is gathered without a specific purpose: information about how long the search took, where in the world it was initiated, what kind of computer and what kind of browser was used, what links (if any) were followed and for how long, as well as information that might be logged that we are not even aware of.

Submitting a query to Google's search engine provides a set of results but also changes the data available on which to base these results in the future. However, the opacity of what and how our actions might be changing is troubling for many: when I use Google Search, what is being taken into account? For instance, say a user sits at the same computer, with the same IP address, in the same location and searched the same search term repeatedly over the course of an hour. We would expect the search results to be the same throughout that hour, outside of instances where a breaking news story might concern one of the search terms. Even if the user receives the same ten search results, the meaning of that search has changed with each search; even if we try and fix or limit our variables, the actions of searching feed and change the system as a whole. Every query and every user's response to the results is part of a very literal instantiation of Gadamer's understanding as a fusion of horizons. Through the aggregation of information drawn from each user, the production of search engine results is simultaneously computationally rigid and organically shifting.

Derrida, in the essay "Typewriter Ribbon: Limited Ink (2)", outlines a position that resonates with the complexities outlined above. He writes,

> Will we one day be able, and in a single gesture, to join the thinking of the event to the thinking of the machine? Will we be able to think, what is called thinking, at one and the same time, *both* what is happening (we call that an event) *and* the calculable programming of an automatic repetition (we call that a machine)? For that, it would be necessary in the future (but there will be no future except on this condition) to think *both* the event *and* the machine as two compatible or even indissociable concepts. Today they appear to us to be antinomic. (72)

The formulation of a search query and using it to query Google's index of the web is far from a living dialogue or dialectic, but it also does not express Plato's notion of fixed language, inscribed and final. The two concepts, the singular event and the repeatability of the machine, converge together in the search process. Derrida argues that such a convergence would represent "a new logic, an unheard-of conceptual form. In truth, against the background and at the horizon of our present possibilities, this new figure would resemble a monster" (73). Derrida's coincidental reprise of Gadamer's notion of converging horizons is quite fortuitous. The two notions of understanding broach one another: the hermeneutics of a static text and the dialectic of spoken communication. Perhaps suggesting that search engines represent Derrida's new logic is an oversimplification of his meaning. In many ways, the monstrous logic that Derrida invokes is, I think, meant to be somewhat impossible. However, it does encapsulate a number of this chapter's concerns about the process of asking questions, defined in the broadest of ways. This monstrous logic recurs throughout this thesis, when addressing algorithmic aggregations and

later discussions of machine learning neural networks; it also encapsulates the tension within how to define a search engine query: queries are part of a long tradition of enquiry, yet technologically distinct from questions.

Conclusion

This chapter has provided a survey of important technical aspects that underpin how search engines function. Outlining scraping and indexing showed that cultural attitudes, as well as market dominance, have an impact on all levels of search engine technology, including even certain technical aspects, such as web crawlers, that might seem too basic to be influenced by such values. Outlining Google's PageRank algorithm explained how the web has grown to be dominated by established institutions and how the authority of such a hierarchy has, in turn, reinforced Google's own hegemony. Addressing query formulation demonstrated how search engine queries are not treated as properly formulated questions, rather, they are a process of iterating citations of language in which their answer might be written. Challenging Groys' provocation that Google establishes a context of words without grammar enabled an analysis that foregrounded the new kinds of grammar that *are* in effect online. Effective searching develops from a deep awareness, conscious or not, of the language patterns that structure the kind of source a user wishes to find. However, the query transaction log analysis showed that, although individuals use search engines for a wide range of activities, their queries are often short and lack specific detail. This establishes a wider concern of this thesis in which Google's responses to such queries rely on successfully predicting the context and intentions of a user, which are dependent on widespread collection and aggregation of data. This issue also contextualises the wider projects of Alphabet, as will be

developed in later chapters, in particular chapter five, which outlines the economic backbone of Google in regard to advertising. Discussions of Plato and Gadamer foregrounded the similarities that queries have within a longer discourse of enquiry, a theme that will continue to be elaborated in the following chapter. Finally, an analysis of algorithmic logic, in which constant and automated feedback plays a role in shifting Google's results every time a search takes place, emphasised the differences between digital and non-digital writing. Unlike in Socrates' reply to Phaedrus that writing remains "solemnly silent" (Phaedrus 275 d-e), digital written discourse represents a new kind of logic. This kind of logic underlies the relationship between questions and queries, while also gesturing to the wider consequences of an algorithmically led culture, that can be highlighted by Derrida's reflection on joining "the thinking of the event to the thinking of the machine" (72). These broader algorithmic issues are developed in chapters three and four, in terms of their social implications and relationship to wider technological directions of Google, for example their increasing reliance on machine learning neural networks. An understanding of the basic operation of search engines and the significance of queries is important for chapter two, as it engages with the topic of memory and widens the historical context of this thesis. In summary, this chapter establishes the technical groundwork for future chapters, while outlining how these issues are embedded within a cultural and social discourse.

Chapter Two:

A History of Technosocial Memory Practices

Introduction

This chapter addresses the relationship between search engine use and memory. There is a significant amount of popular literature and journalism that make alarmist claims about the impact of ubiquitous computing; the most infamous example is Nicholas Carr's anecdotal approach in his 2008 article "Is Google Making Us Stupid?" and 2010 book, *The Shallows*.[1] Rather than engaging with popular accounts directly, this chapter addresses current research in the field of psychology, which is used as a jumping-off point to draw on a long history of memory practice and its inherent social and philosophical issues. The chapter outline is as follows. First, this chapter discusses the psychological research concerning search engine use and its potential impact upon the information retention of individuals. In doing so, the analysis establishes the importance of technological metaphor in such debates. Second, a phenomenological approach to search engine use is outlined, in order to establish some key aspects of online enquiry. Third, a longer history of technology is established, which argues that personal memory has always functioned dynamically by relying on external technology and other individuals. There are, however,

[1] Such writing is complete with a number of subgenres, including: generalisation regarding technology's impact on social change (Andrew Keen's *The Internet is Not the Answer*, 2015); self-help (Cal Newport's *Deep Work: Rules for Focused Success in a Distracted World*, 2016); apocalyptic visions (Gerd Leonhard's *Technology vs. Humanity: The Coming Clash Between Man and Machine,* 2016); and attitudes of determinism (Kevin Kelly's *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, 2016).

particular moments of profound change such as the birth of literacy and its impact on oral culture, an event that is often used as a parallel for our contemporary shift towards an increasingly digital culture. Therefore, engaging with the shift from orality to literacy enhances current debates surrounding the influence of Google and the web. Drawing on a longer history demonstrates that metaphors and mental models of complex processes, such as memory, draw from technological allegories and have significant effects on the discourse surrounding these processes. Fourth, I argue that many of the debates regarding knowledge and memory can be framed within the structure of either a Platonic or Aristotelian epistemology. These two approaches, first outlined in the context of a shift from orality to literacy, are developed throughout this chapter in other technological contexts, leading up to the present. Drawing from the oppositional approaches of Plato and Aristotle in regard to memory also offers a context for later debates, particularly in relation to the various ways in which theorists, programmers and cultural critics have defined what they see as the fundamental characteristics of search engines. Finally, this chapter addresses the set of technosocial memory practices known as the *Ars Memoria*, which had significant use between 400 BCE and 1600 CE and a form of which is still practised by contemporary competitive memory athletes. The history of the *Ars Memoria* illustrates the tension between Platonic and Aristotelian attitudes, the relationship between logic and mysticism, and the rise of modern science. This historical and theoretical background is used to provide depth to the original question regarding the impact of search engine use on memory. The chapter concludes by underlining the importance of understanding search engines as a socially embedded technosocial practice, of a kind that has a long history, that carries certain implicit social and philosophical attitudes.

1.0 Google at Our Fingertips

Sparrow et al., in their 2011 study “Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips”, report the findings of four experiments carried out to explore the potential effects of search engines on memory recall. Each experiment tested a different aspect of search engine use; specifically, would participants think about computers when faced with question? (study one), would participants’ memories be affected by whether or not they thought information would be available in the future? (studies two and three), would knowing where to find a piece of information have a relationship to the memory of that information (study four). Their results suggested that

> when faced with difficult questions, people are primed to think about computers and that when people expect to have future access to information, they have lower rates of recall of the information itself and enhanced recall instead for where to access it. The Internet has become a primary form of external or transactive memory, where information is stored collectively outside ourselves. (776)

Daniel Wegner, one of the researchers in the above study, first established the term “transactive memory” in a 1985 article “Cognitive Interdependence in Close Relationships”, written with Toni Giuliano and Paula Hertel. Wegner et al. found that couples interdependently relied on the memories of one another and through the unintentional practice a number of techniques, cueing and re-cueing for example, couples could arrive at information that neither would have successfully recalled

alone. Later research (reported in Wegner et al. “Transactive Memory in Close Relationships”) built on the previous findings that couples rely on the memories of one another, to present evidence that couples unconsciously establish a division of labour in regard to the types of things each remembers and forgets. Wegner et al. argued that this impacted on future abilities to recall certain types information. The central study of “Transactive Memory in Close Relationships” tested the recall of established couples and impromptu couples (pairs comprised of strangers). These two types of couples were split into two further groups, the first were assigned particular memory structures, for example, “1 partner should remember food items, another should remember history items” (923), the second were given no such instructions. The results of the study found that “Memory performance of the natural pairs was better than that of impromptu pairs without assigned structure, whereas the performance of natural pairs was inferior to that of impromptu pairs when structure [sic] was assigned” (923). The second part of this finding, that established couples performed worse than impromptu couples when they were instructed to follow specific memory roles, is key to Wegner’s definition of transactive memory. Over time, couples establish implicit and unconscious memory roles that Wenger et al. argue influence the capacity to form memories as well as recall them.

With this theoretical background in mind, the assertion of Sparrow et al., that search engine use represents an example of transactional memory, is a bold claim. One of their four experiments ran as follows:

> Participants were tested in a 2 by 2 between-subject experiment by reading 40 memorable trivia statements of the type that one would look up online (both of the new information variety, e.g., “An ostrich’s eye is bigger than its

> brain," and information that may be remembered generally, but not in specific detail, e.g., "The space shuttle Columbia disintegrated during re-entry over Texas in Feb. 2003."). They then typed them into the computer to ensure attention (and also to provide a more generous test of memory). Half the participants believed the computer would save what was typed; half believed the item would be erased. In addition, half of the participants in each of the saved and erased conditions were asked explicitly to try to remember the information. After the reading and typing task, participants wrote down as many of the statements as they could remember. (776)

They found that "those who believed that the computer erased what they typed had the best recall […] compared with those who believed the computer would be their memory source" (777). In addition, there was no statistically significant difference between the groups that were explicitly asked to try to remember the information. Reflecting on the experiment in 2013, Sparrow and Chatman conclude that the results "suggests this is not a conscious decision people are making" (276), which is consistent with previous findings of Wegner et al. in regard to the roles that couples had established but were not aware of.

Another key point of discussion in their 2013 paper was the differences between the particular affordances of search engines as transactive resources and other kinds of transactive memory. In particular, they focused on the process whereby individuals do not remember a piece of information but can remember where this information can be found. Sparrow and Chatman's findings were that

> when we gave people highly memorable trivia (“An ostrich’s eye is bigger than its brain”) and very unmemorable places the information they were typing would be stored (things), they tended to remember one or the other. If they remembered the trivia itself, they did not remember where to find it and vice versa. And aside from remembering neither, they were most likely to remember where to find the information. (276)

Highly memorable trivia was, in fact, less memorable than the generically interchangeable folder names, such as “FACTS, DATA, INFO, NAMES, ITEMS, or POINTS [capitalisation used in the original study]” (2011). Sparrow and Chatman (2013) highlight how this kind of memory practice has deep historical roots, but that, given the ubiquity of search engine access, particularly its increasing access on mobile devices, prioritizing where information might be found over the information itself may be much more useful. Sparrow and Chatman also challenge the belief that such behaviours have a negative impact on memory, arguing for three beneficial aspects drawn from their findings. First, they argue that when individuals “repeatedly visit similar sources of information that arise from online searches” the resulting information may be “better learned in the long run” but also that individuals “will selectively learn the information that is relevant in that informational context” (277). This context-dependent attitude towards information access echoes Google’s own notion of relevance (outlined in chapter three). Second, they reflect on a related study to their 2011 experiment, outlined above, which gave participants a problem-solving activity. They found that in comparing the participants who thought the information was erased with the participants who were informed that it would be

accessible, the latter group recalled fewer pieces of information but were better at solving problems.

> Participants who memorized the details, who believed the problem would be inaccessible to them later, did in fact recall more details. But they solved fewer problems. In addition, across conditions, the number of details remembered negatively predicted the number of insight problems successfully solved. This evidence suggests that offloading the remembering of details onto the Internet as a transactive memory partner will in fact aid creative problem solving. (278)

This result provides more nuance to the conclusion that search engine use impacts memory. Finally, Sparrow and Chatman reflect on a very specific aspect of web search engine access: choosing a result. Evidence from their studies demonstrated that when participants selected a search engine query result, their sense of agency and control increased. This was the case even when participants choose either the top results, or their “choice” actually led to an unrelated result. They found that the more agency participants felt they had in excursed in finding a piece of information, the more inclined they were to be critical of that information. Which, as they argue, “is pretty much the opposite of what most people believe, which is that we, especially if we are children, believe everything we read online without question” (280). It is significant that the structure of search engine enquiry forces users to make a choice, which distinguishes it from other modes of transactive memory. In forms of transactive memory that rely on other individuals, the recalled information is not necessarily scrutinized by the apparent self-doubt directed at search engine results.

Sparrow and Chatman argue that: “Because it feels like we are in charge of our web searches, even if we are not as agentic as we believe, the lack of transparency, paradoxically according to our preliminary research on agency and evaluation, results in web users being more critical in the evaluation of what they read online” (281). None of the authors have published any more substantial findings on the relationship between perceived agency and critical scrutiny since 2013; therefore, the findings of these preliminary studies should not be overstated, or taken to directly reflect contemporary digital practices. The final chapter of this thesis, which highlights the extent to which fake news stories are shared by individuals, provides some evidence that individuals can be very uncritical of sources, although this is, of course, bound-up with several political complexities that will be discussed later.

1.1 Can Google Really Provide Transactive Memory?

Although no literature disputes the methodology or results of Sparrow et al. and Wegner et al., there are those who have been critical of the interpretative structures used to draw conclusions. Bryce Huebner’s 2016 review of transactive memory, in which he also includes Andy Clark and David Chalmers’ extended mind thesis of 1998 and subsequent literature, draws out the theoretical factors involved in making claims that the Internet could be used as a site of memory or cognition. Huebner agrees that there is a “broad and expanding consensus that we often exploit the physical and social structure of our world when we expect that the information we need will be available when we look for it again” (49). However, Huebner argues that defining such activities as memory requires a specific attitude towards the relationship between cognition, agency, and memory. In particular, Huebner draws

from John Sutton's *Philosophy and Memory Traces: Descartes to Connectionism*, a history of philosophical accounts of memory, which argues that attitudes towards memory can be categorized into two camps: archival and constructive. Huebner sides with the constructivists and contends that memories represent "skeletal representations" (59) with which individuals flesh out with tacit and general knowledge at the time of remembering. Huebner draws on a body of research concerning false memories and eye-witness accounts, which address the frequency of false details that an individual supposedly remembers. Important to Huebner's argument is that these false memories are often unrelated to the event and, therefore, have been constructed, rather than misremembered. The archival approach, that sees memories as far more substantial and static, does not account for such evidence. Huebner argues that using a constructive approach

> to transactive memory systems suggests that Wegner was wrong to claim that our frequent and pervasive use of Google searches and iPhones is sufficient to establish the existence of novel transactive memory systems. In these cases, the flow of information is unidirectional and exploitative. In these cases, we find a person who uses the informational resources and who encounters information that is structured in a way that makes it a target for exploitation. But exploitation is the paradigmatic relation that obtains between a person and the tools that she uses. (64)

A hypothesis that Huebner mentions in passing, but that is key later in this chapter, is that the increasing use of technology such as Google and smartphones might predispose individuals to model their approach to memory as "archival" as this "sits

comfortably with an everyday understanding about how digital computers work" (56-57). Douwe Draaisma argues in *Metaphors of Memory: A History of Ideas About the Mind* that the line between metaphor and usage is significantly blurred:

> artificial memories have not only supported, relived and occasionally replaced natural memory, but they have also shaped our views of remembering and forgetting. Over the centuries memory aids provided the terms and concepts with which we have reflected on our own memory. We have 'impressions', as if memory were a block of sealing-wax into which a signet ring is pressed. Some events are 'etched' on our memory, as if the memory itself were a surface for engraving upon. What we wish to retain we have to 'imprint'; what we have forgotten is 'erased'. We say of people with an exceptionally powerful visual recall […] that they have a 'photographic memory' [and therefore our] views of the operations of memory are fuelled by the procedures and techniques we have invented for the preservation and reproduction of information. (3)

Draaisma argues that these metaphors are not simply used as models, which come and go, but structures that have caused historical effects on how we conceive of the mind and consequently the technologies and practices that continue shift such a definition.

Technologies are formed as consequences of social attitudes towards cognition and memory and, in turn, shape those attitudes as they become socially embedded. Huebner's above contention is that the use of search engines and smartphones as memory prostheses has encouraged the perspective that these

activities replicate a model of the mind, that a hard drive memory measured in gigabytes might be comparable to human recollection. Such attentiveness to the potential biases of our own milieu is valuable. When Draaisma goes on to say that “The history of memory is a little like a tour of the depositories of a technology museum” (3), he is not arguing that similarities of expression exist, but that technologies, metaphors and practices are deeply enmeshed. Arguing whether or not search engines represent a form of transactive memory depends on a particular definition of memory, which itself is based upon the attitudes and practices informed by ubiquitous computing. However, with this cautionary perspective in mind, I contend that Huebner’s definition of search engine use as “unidirectional and exploitative” (64) is not an accurate portrayal of how search engines function. A brief phenomenological example demonstrates that search engines can, in some circumstances, be considered as a form of transactive memory.

2.0 The Phenomenology of a Search

Due to the extent to which search results change between contexts and throughout time, providing examples from which to generalise is difficult, if not an impossible task. In addition, as established in chapter one, individuals use a wide range of search behaviours. The key to understanding algorithmically governed phenomena does not lie in trying to fix a representative sample of results for study, as this conceals the dynamic nature of the object of study. Such a comparative analysis can be useful for some aspects, for example demonstrating variation between contexts, and is an approach taken in chapter four. Methodologies that focus on a smaller scale can prove useful ways to highlight technological characteristics, even if these individual examples cannot claim to be generalizable. For this reason, this section's approach draws from phenomenological methods and the thick descriptive accounts of Clifford Geertz,[2] to demonstrate the transactional dynamics of a single search session.

The search session in question is the process of trying to find out the name and artist of a song stuck in my head. I have very little information to work from; I do not know the name of the artist, whether it is a new or an old song and whether or not it has lyrics. I do, however, have another non-linguistic cue to work from. The sound of the synth from the song is paired in my mind with an image of a sun-drenched car, filled with cigarette smoke, circling around an actor, the name of whom I cannot remember either. I hazard a guess that the actor I can picture in my mind starred in a recent feature film about Facebook. So, I start Googling, I search [film about facebook], the top hit of which is *The Social Network* (fig. 1).

[2] See Geertz's "Thick Description: Toward an Interpretive Theory of Culture" (1977).

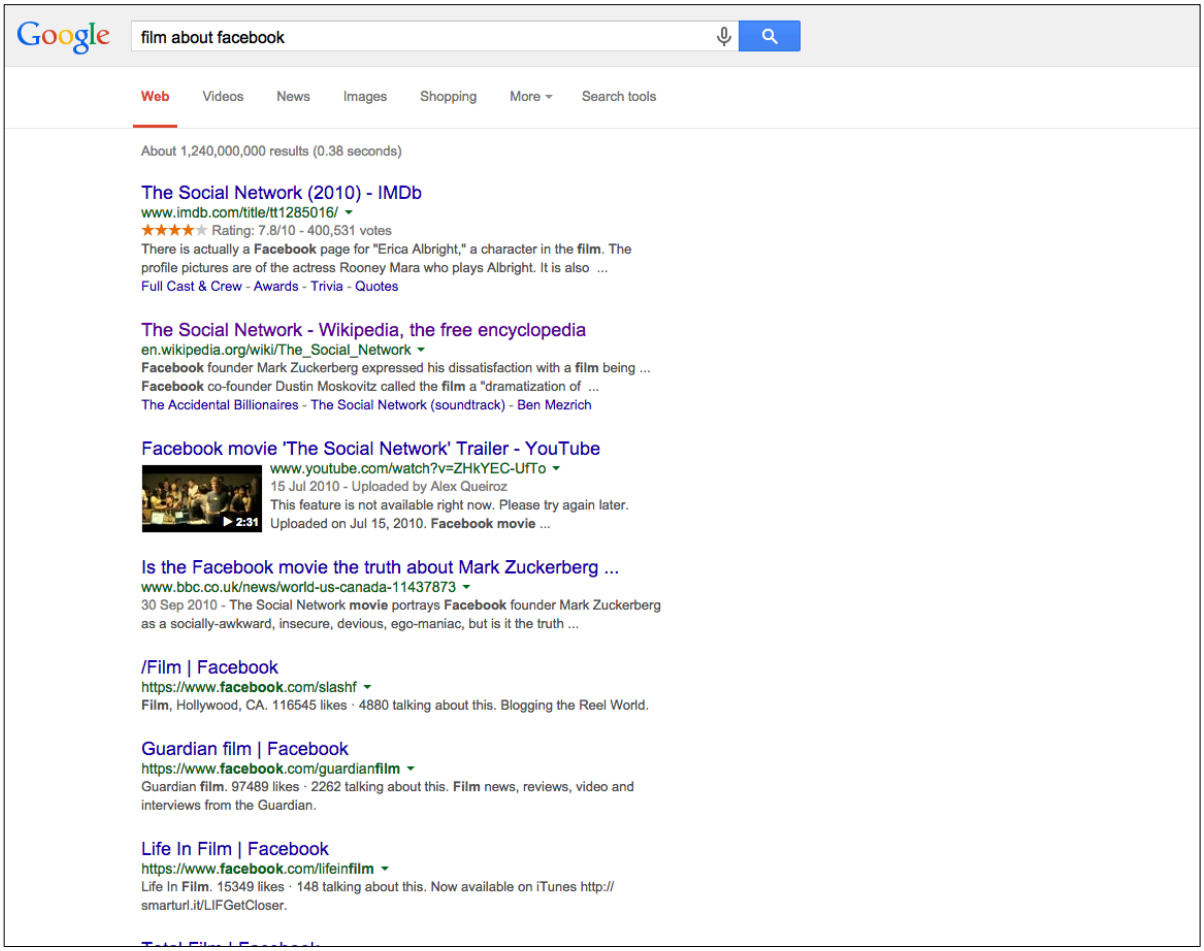


Figure 1. A Google search for the query [film about facebook]. Search performed 23/03/2015.

I click back into the search box in order to refine my search query and a list of suggested searches pops up in an autocomplete dropdown list (fig. 2)

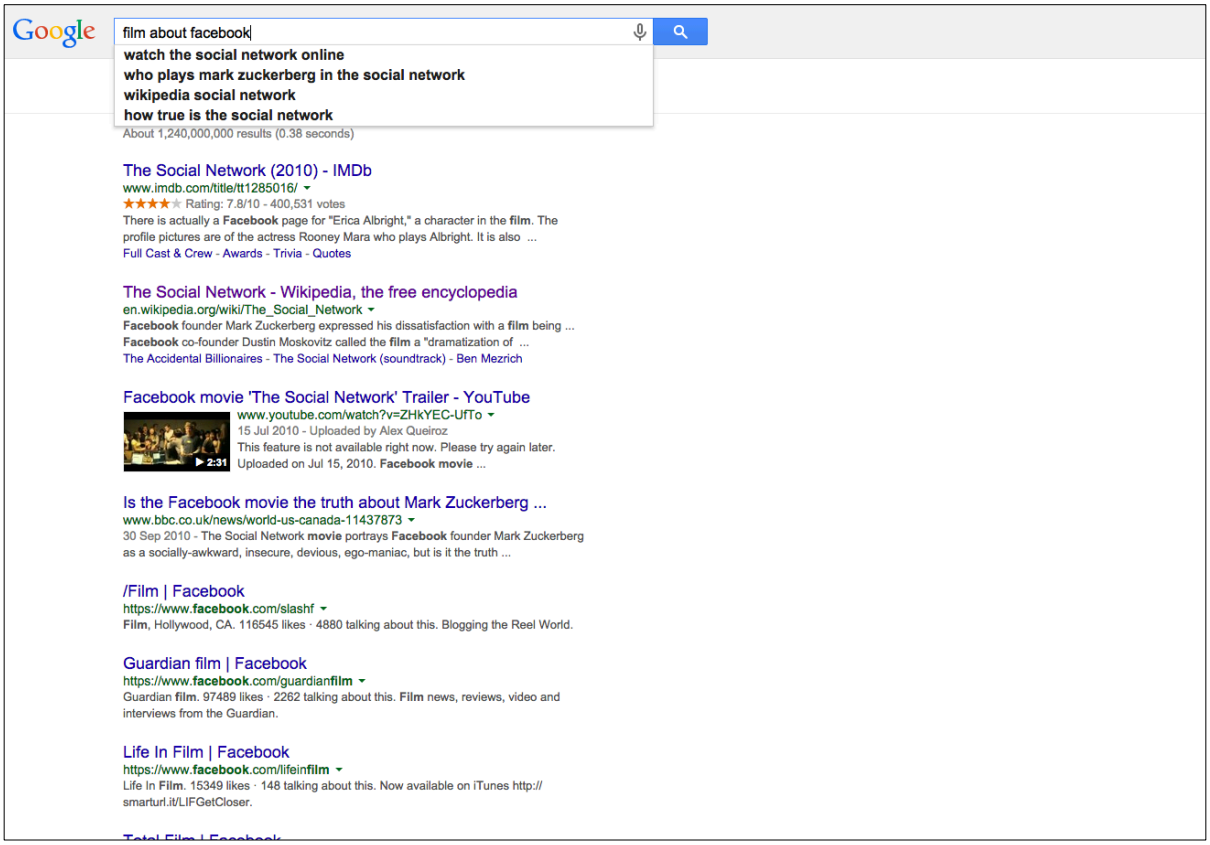


Figure 2. A Google search for the query [films about facebook] with Autocomplete suggestions. Search performed 23/03/2015.

The Autocomplete function, which is the focus of chapter three, provides a user with a range of related search queries. Rather than functioning as a simple predictor of the end of a user's stem, for example in [who plays mark zuckerberg in the …], as the technology was originally conceived, Google informs the user of other search patterns, questions, or relations, in a process of cueing that, I argue, reflects its role as a source of transactive memory. I choose the second down, [who plays mark zuckerberg in the social network] (fig. 3).

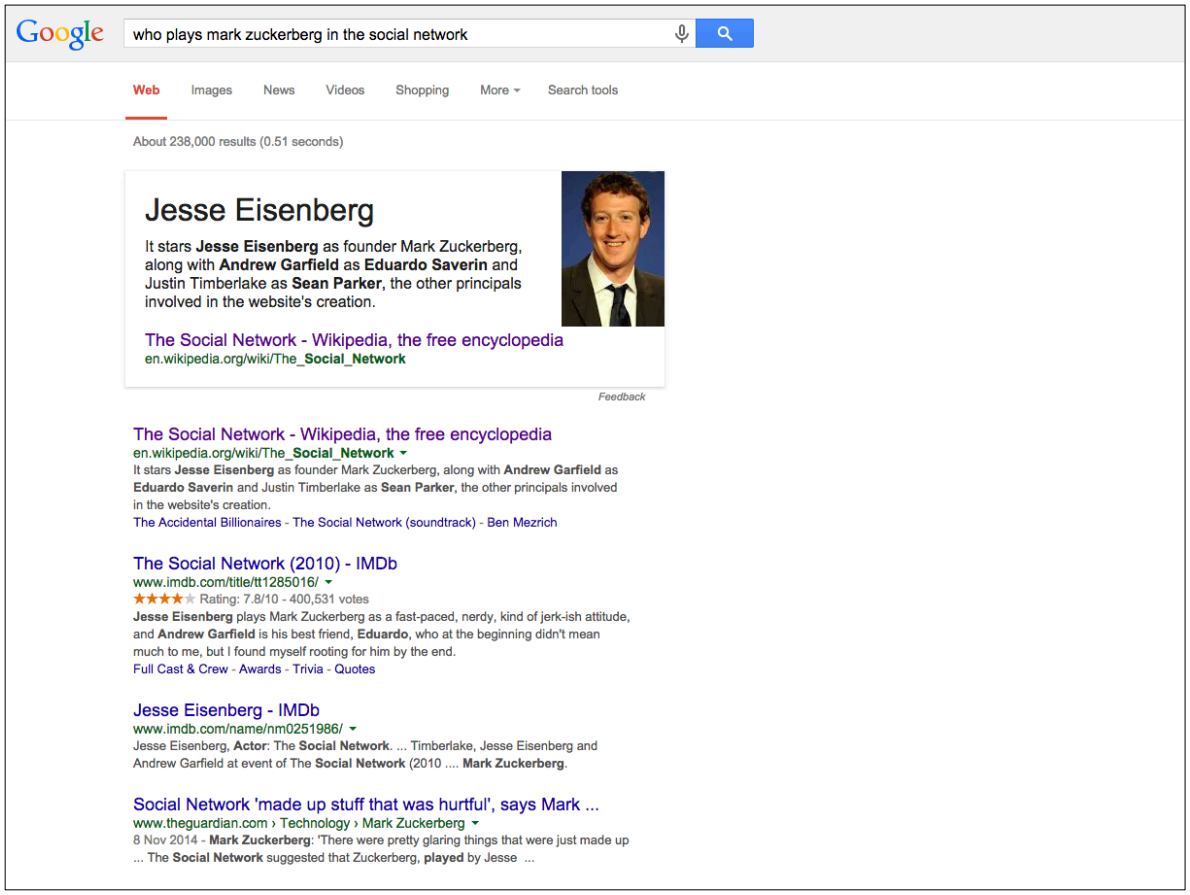


Figure 3. A Google search for the query [who plays mark zuckerberg in the social network]. Search performed 23/03/2015.

In another example of its agentic capacity, this suggested search is answered using Google's semantic network, knowledge graph. This knowledge panel, drawn from Wikipedia, informs me that the lead actor of *The Social Network* is Jessie Eisenberg. I follow the second link provided, Eisenberg's Internet Movie Database (IMDB) page, read through his filmography and feel that none of them is the film I am trying to find.

I return to Google and search [who looks like Jesse Eisenberg] (fig. 4) and am informed through the title of a result that Michael Cera and Andy Samberg represent potential leads. Again, Google is supplying the user with information without the need to follow any of the search results.

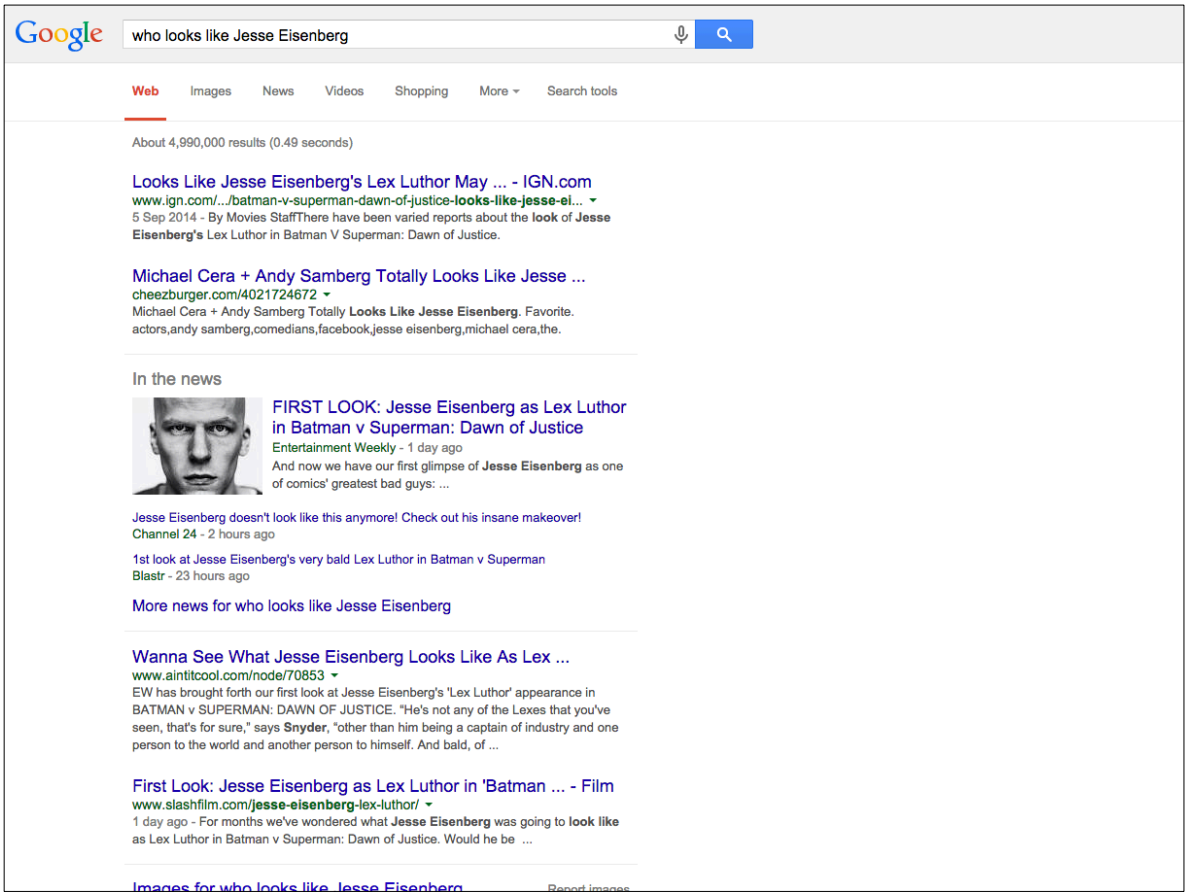


Figure 4. A Google search for the query [who looks like Jessie Eisenberg]. Search performed 23/03/2015.

I start with Michael Cera, not only because he is listed first so I assume that he might be a closer doppelgänger than Samberg, but also because I know that I have seen more of Cera's films. I switch to Google image search in the hope that a promotional poster or film still might be recognisable and search [michael cera films] (fig. 5).

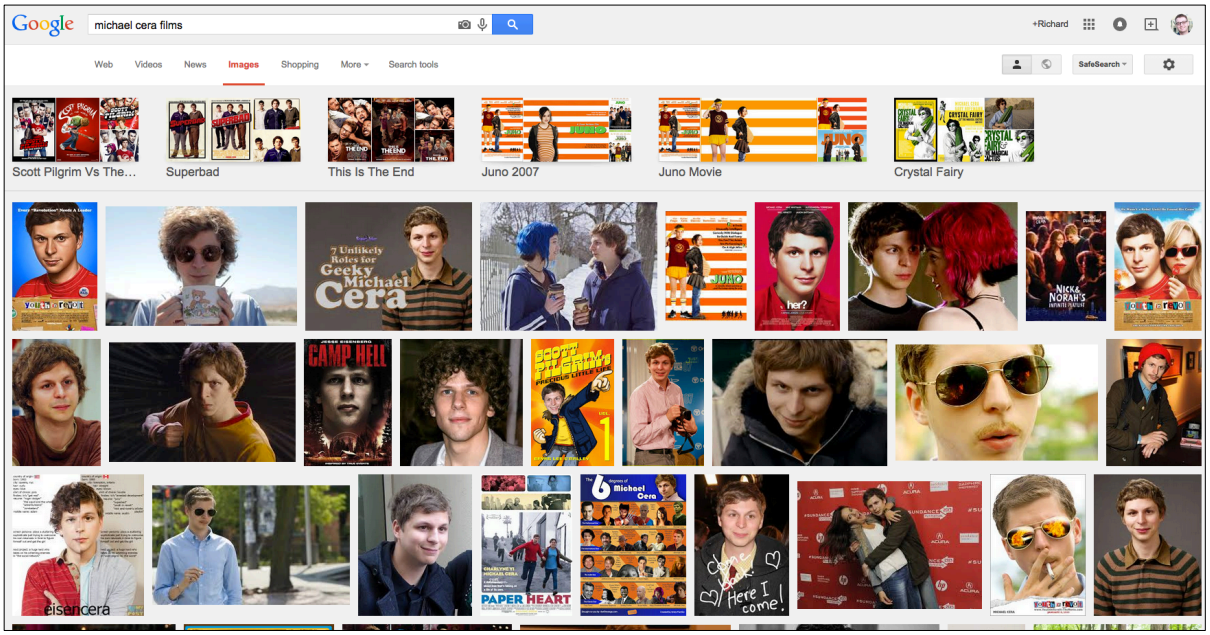


Figure 5. A Google Image Search for the query [michael cera films]. Search performed 23/03/2015.

None of the images relate to the picture I have in my head, consequently, I refine my query, searching [michael cera films car] (fig. 6).

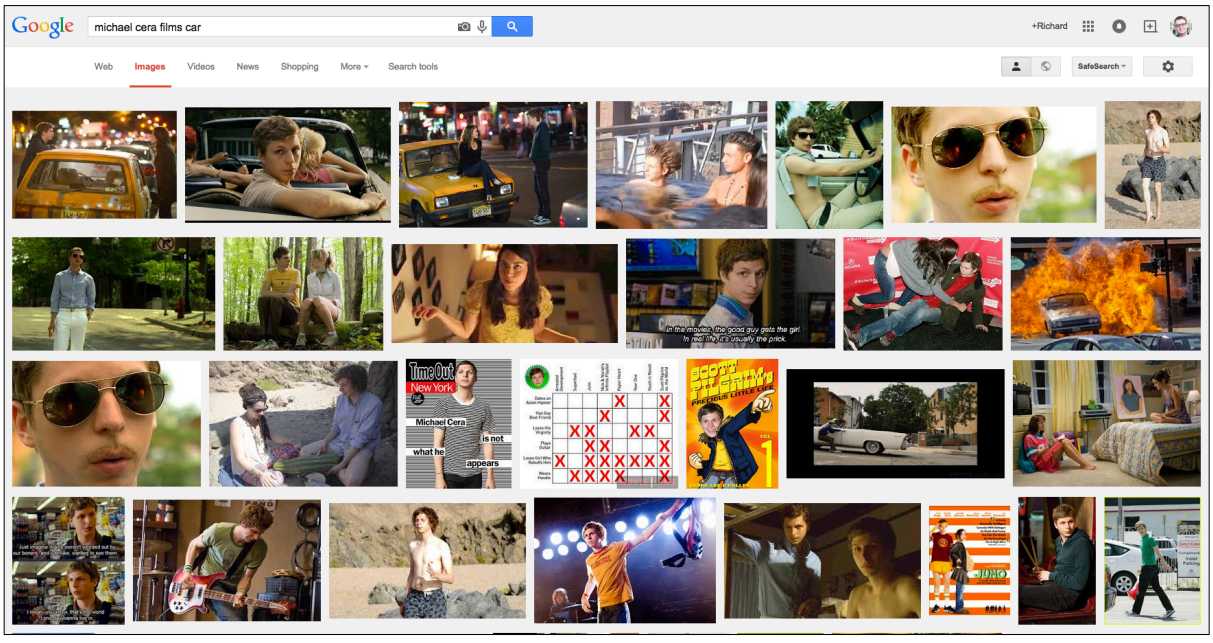


Figure 6. A Google Image Search for the query [michael cera films car]. Search performed 23/03/2015.

Although I do not have any detailed memory of what kind of car I am looking for, my memory of it drenched in sunlight does not fit the context of any of the images provided. Therefore, I change tack and search [michael cera film smoking] (fig. 7)

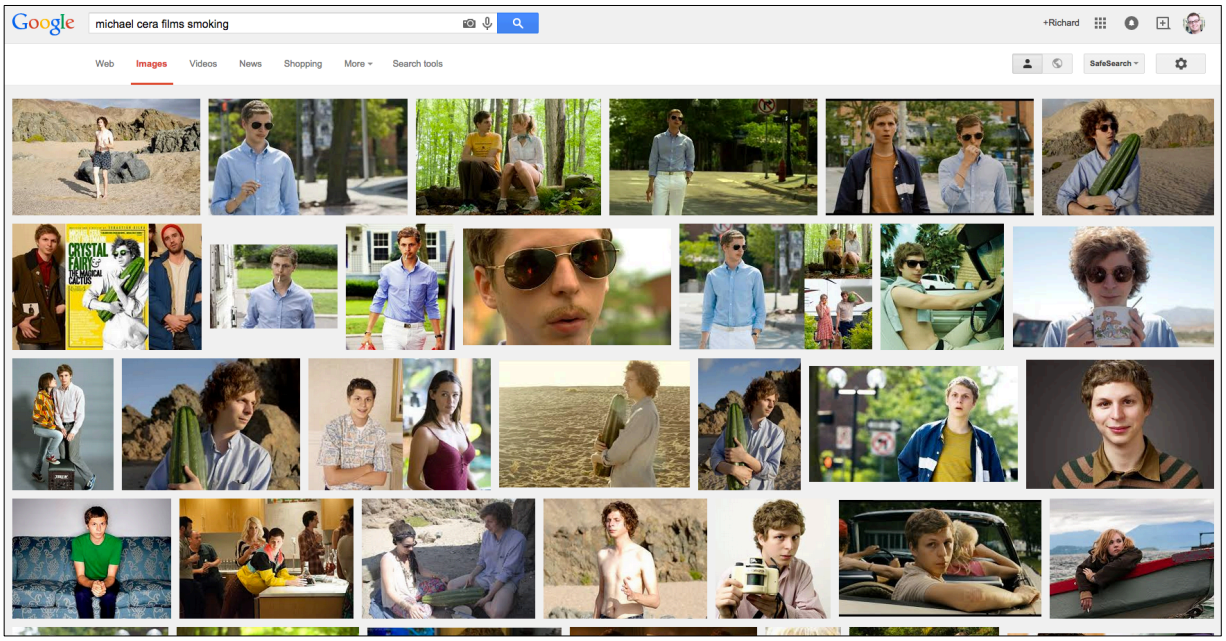


Figure 7. A Google Image Search for the query [michael cera films smoking]. Search performed 23/03/2015.

The grid of images presents an answer in aggregate; none of the pictures individually represent the image I have in my mind, instead their aesthetic gestalt chimes with my memory. Clicking on a particular thumbnail, Google informs me that the picture is taken from a review written by Alex Godfrey in *The Guardian* for the film *Crystal Fairy* (fig. 8).

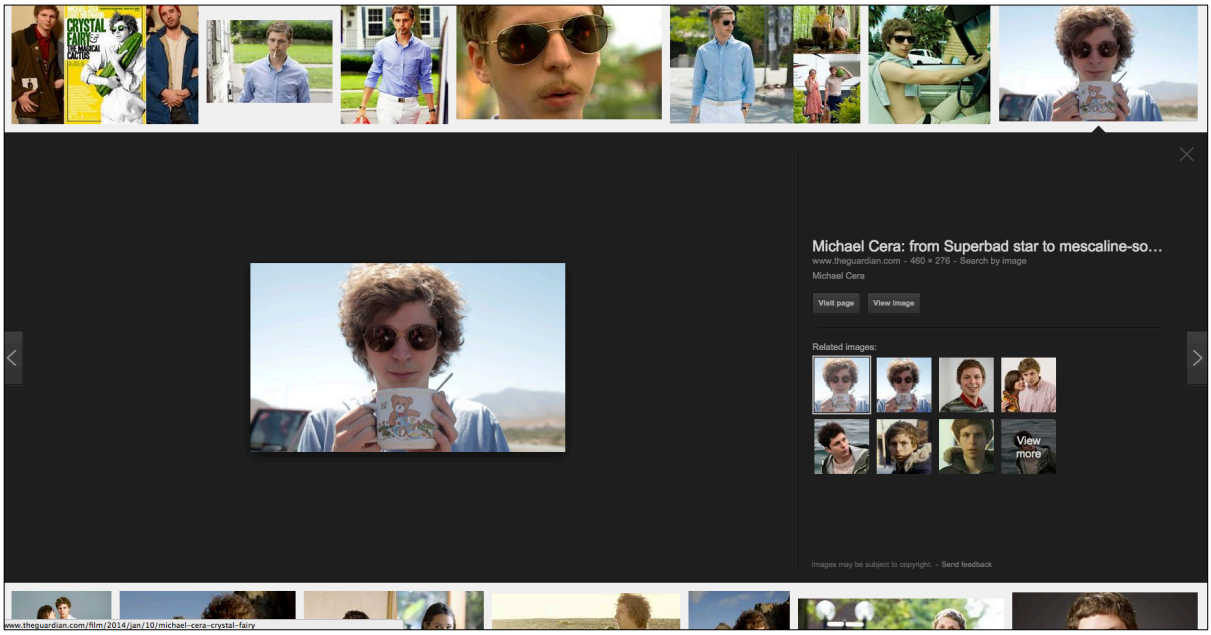


Figure 8. A particular Google Image result for the query [michael cera films smoking] showing an image used in the article “Michael Cera: From Superbad Star to Mescaline-Soaked Drug Adventurer” (Godfrey), with a link to the webpage of *The Guardian.* Search performed 23/03/2015.

Coincidentally, the context of the image, the other thumbnails in fig. 7, plays a more significant role than I realised at the time. It was only when reviewing the screenshots for this chapter that I noticed the same image is the second result in the earlier search (fig. 5) of [michael cera films].

Using the information provided by the image search, I switch back from Google’s image search to Google’s main web search and enter the query [Crystal Fairy] (fig. 9).

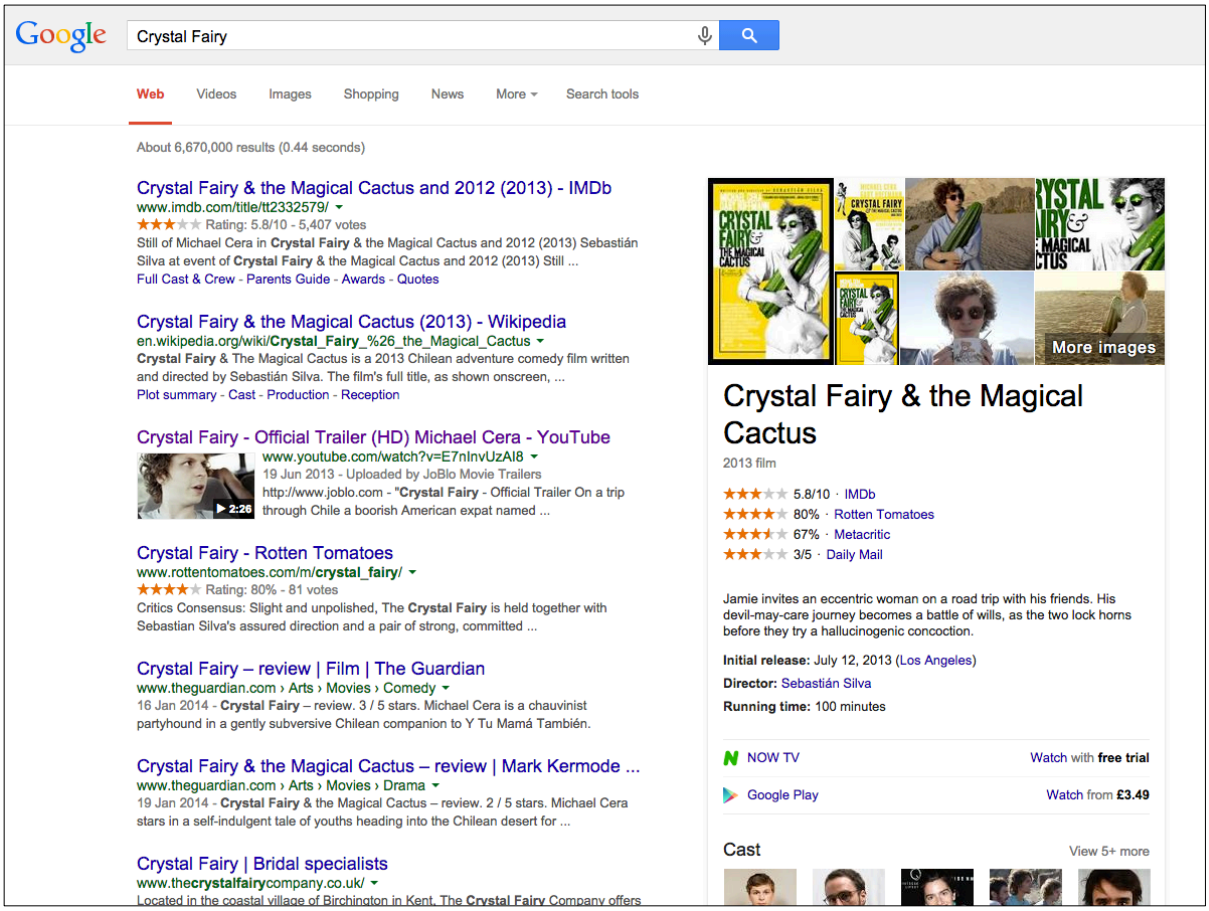


Figure 9. A Google search for the query [Crystal Fairy] Search performed 23/03/2015.

From here I follow the result for the trailer on YouTube (fig. 10). The song I am trying to find soundtracks the opening of the trailer. I scroll down to the video's description hoping the song is listed, but it is not. Aware of the convention that users often answer questions in the comments, I continue scrolling until I get to the comment: "whats the name of the first song playing", which is answered with the comment: "Bongo bong - Manu Chao" (fig. 10).

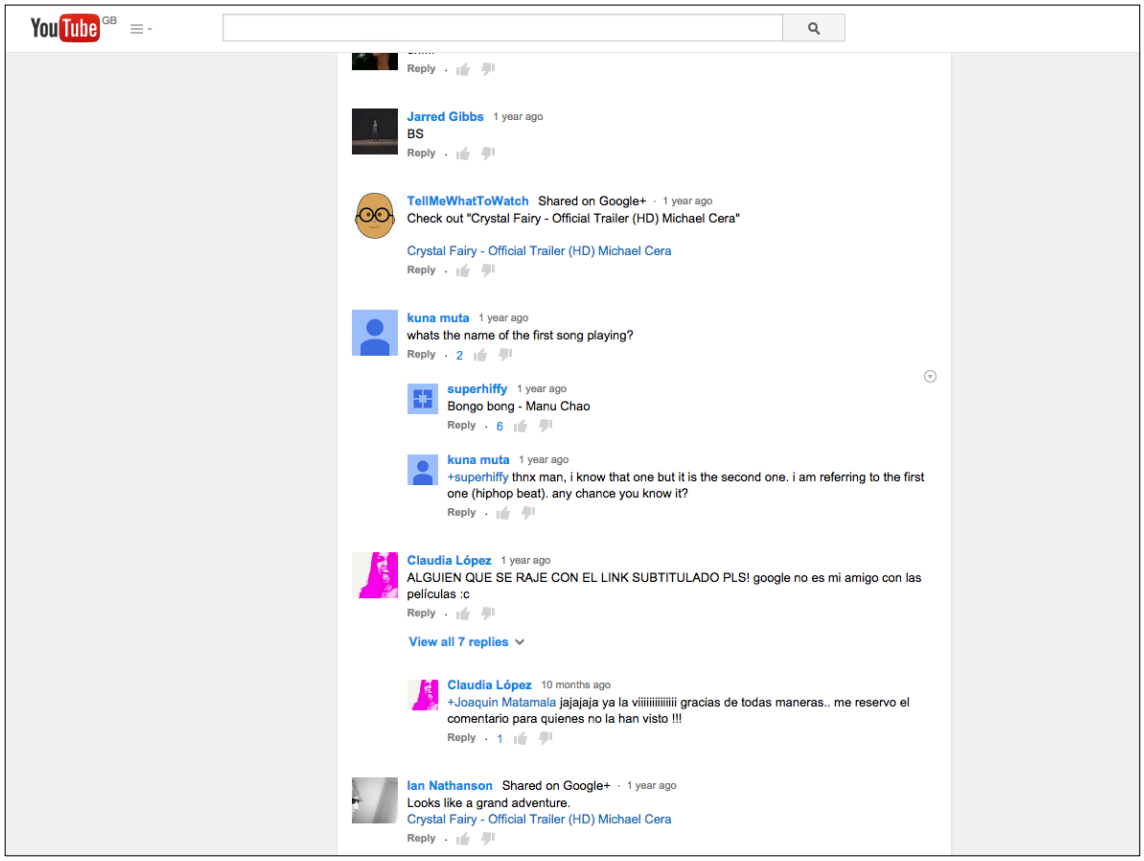


Figure 10. YouTube comment by “superhiffy” on “Crystal Fairy – Official Trailer” Accessed 23/03/2015.

My question has been answered; the whole process took no longer than a couple of minutes.

The example continued to anecdotally support the conclusions of Sparrow et al. (2011), as, on a number of occasions after the search process outlined above I wanted to listen to the song again. However, I still could not remember the name of the song or artist although I could remember the search pattern that led me to the trailer. Following a shortened version of my previous method, the query [michael cera drug film] led me to the YouTube video, down to the comment listed in fig. 10 and to the query [Bongo bong - Manu Chao], which led to a full version of the song on YouTube. In addition, I have followed this path perhaps ten times since; I

continue to remember the associative pathway but not any of the information regarding the artist or song title. I argue that this personal example presents an interaction with Google's search engine that goes beyond Huebner's description of Google as a "unidirectional and exploitative" (64) tool. Not only did Google provide cues and suggestions, but it did so based on the search patterns of other users. Performing the searches and following the results provides data that reinforces these associative pathways for myself and other users. Over time if other users followed the same pattern the links followed would rise in the results, additional autocomplete suggestions would become more specific and my end result, the video for "Bongo bong" would possibly end up becoming a result for a query about Jessie Eisenberg, even though he was not in the film. Instead, the results that do appear provide evidence that this association is, unsurprisingly, not the main reason people are searching for Jessie Eisenberg, or even Michael Cera for that matter.

Google do, however, provide personal autocomplete suggestions based on exact previous searches of each individual user. For example, when starting my query with [m], after having searched [michael cera drug film] the autocomplete provides two previous searches, verbatim, and two associated queries (Mary Elizabeth Winstead starred alongside Michael Cera in the film *Scott Pilgrim vs. the World*) (see fig. 11).

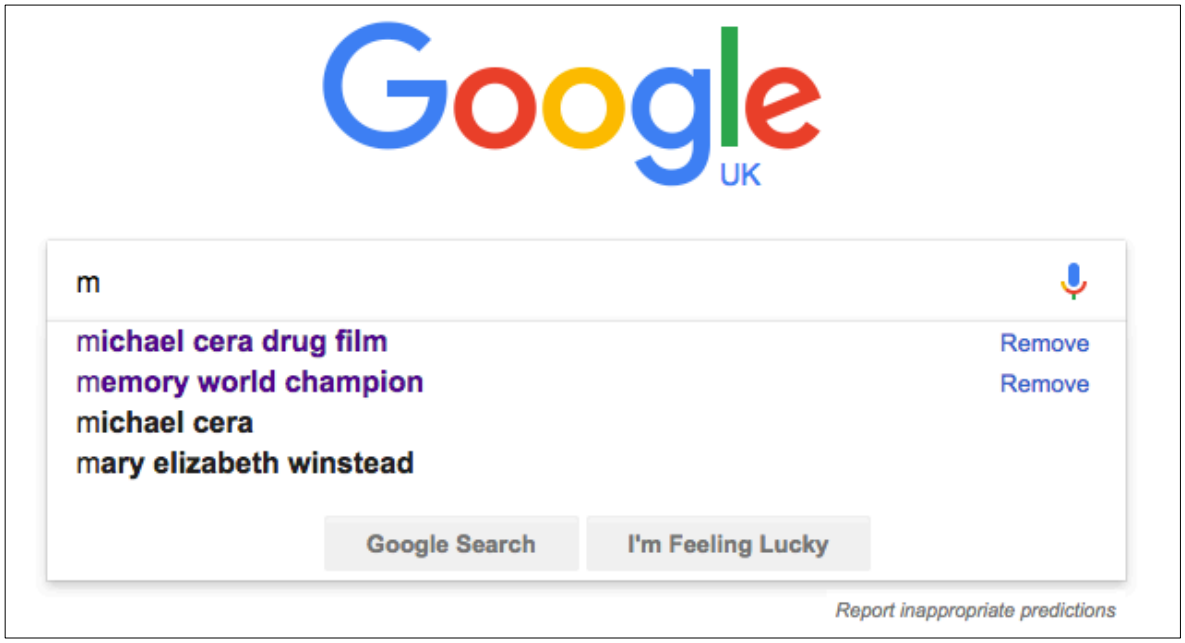


Figure 11. Autocomplete suggestions for a query stem beginning with [m], showing two personalised suggestions. Search performed 30/07/2017.

The autocomplete function will be discussed in greater depth in chapter three, for now this example is a demonstration of Google's bidirectional agency. Users are not in control of which previous searches are included; they can remove individual queries, but the various criteria used are hidden. This feeds into the personalization search results, with reminders below specific search results detailing the frequency of visits and date of the last visit (see fig. 12) as well as other features of personalization that will be discussed throughout this thesis. With these examples in mind, this chapter returns to the opening question regarding the interplay between technology and memory.

Figure 12. An example of a Google Search result that details previous visits.

Screenshot taken 29/11/2017.

3.0 Is the Shift from Oral Culture to Literacy a Useful Historical Precedent?

A number of media theorists have drawn a parallel between the shift from orality to literacy and the contemporary shift from a print-based literacy to a form of electronic literacy. Building on the work of Walter Ong,[3] which demonstrated the impact of writing and the alphabetic system upon oral cultures, theorists have claimed that the digital era produces a comparably fundamental shift. Gregory Ulmer, for instance, coined the neologism “Electracy” (43) (a portmanteau of electronic literacy) building on Derrida’s Grammatology to enhance Ong’s notion of second orality “as a hybrid sharing features of literate and oral practices, to be understood as intervening between and mediating the apparatuses of orality and literacy, distinct from its chronological position as coming after literacy” (Ulmer 163-4). In particular, whereas Ong stresses oral media forms that exist in literate societies, such as the radio, Ulmer reinforces a Foucauldian emphasis on the institutions that gain power in a digitally mediated space. Another theorist who draws a parallel between the birth of literacy and digital communication is Friedrich Kittler, who places the emphasis on the open-ended nature of alphabetic encoding, writing that “for a second time in history, a universal medium of binary numbers is able to encode, to transmit and to store whatever will happen, from writing or counting to imagining or sounding” (“Towards an Ontology of Media” 2-3). Even though the emphasis shifts between theorists, the claim made that discursive epistemes are underpinned by technological reciprocity remains consistent. It is useful to outline the impact that

[3] Ong’s work relies on the research of Eric Havelock (see *Preface to Plato* (1963)), Albert Lord (see *The Singer of Tales* (1960)), and Milman Parry (see *L'épithète Traditionnelle Dans Homère: Essai Sur Un Problème De Style Homérique* (1928)).

literacy had on oral culture for three reasons. First, such a shift can provide a historical grounding to claims about current technology such as Google, and their effects on language and communication. Second, such a discussion provides a way in to outline the epistemological underpinnings regarding the philosophy of memory, through the opposing views of Plato and Aristotle, which will be invaluable for a discussion of the current functions of search engines as memory prosthetics. Finally, the narrative of literacy leads directly on to an important historical precedent for current search engines, the set of memory techniques known as the *Ars Memoria*.

Walter J. Ong's *Orality and Literacy: The Technologizing of the Word* (1982) proposes the argument that literacy marks a significant, if not the first, form of technology that fundamentally alters the way in which humans think. Before, entering into greater detail of Ong's line of argument, particularly in relation to memory, one aspect should be outlined first. That is, for Ong, literacy fundamentally changes the way people think and once someone has learned the ability to write they cannot return to their previous cognitive system. A central example for Ong is that of preliterate texts, known more commonly as oral literature,[4] such as the Homeric epics. In referencing Milman Parry's ground-breaking study of Homeric poetry published in 1928 as *L'Epithete traditionelle dans Homere* (in English: *The Making of Homeric Verse)* Ong states that "Parry's discovery might be put this way: virtually every distinctive feature of Homeric poetry is due to the economy enforced on it by oral methods of composition" (21). Its form and content are enmeshed in a fundamental way. Ong describes how oral texts can never, due to their medium specific nature, be reproduced in exactly the same way twice. Nor can they follow a

[4] Although Ong fervently rejects the term oral literature due to the retroactive use of the word literature, which characterises certain structures and features which are exactly what oral texts *can never* be.

linear structure of written texts. It is not that oral texts *fail* to be linear but instead, that exact sequence has no relevance to oral texts, as Ong explains:

> The singer is not conveying 'information' in our ordinary sense of 'a pipe-line transfer' of data from singer to listener. Basically, the singer is remembering in a curiously public way – remembering not a memorized text, for there is no such thing, nor any verbatim succession of words, but the themes and formulas that he has heard other singers sing. He remembers these always differently, as rhapsodized or stitched together in his own way on this particular occasion for this particular audience. (145-146)

For oral poets, recall represents an awareness of fluid relationships. This covers a range of skills from the knowledge of the length and mutability of stock phases used to complete the meter of each phrase, for instance, dactylic hexameter, up to a macro understanding of themes and tales that might lead into others or create possibility. To use Ong's terminology, this knowledge develops from interpersonal experience embedded in the human lifeworld. There is not a set of principles taught by oral poets or a correct way that a text should flow but an understanding of context, experience of manipulations, and open-ended understanding of memory as praxis. This understanding of how oral texts functioned also applied to everyday experience. As Havelock asserts "in the Greek situation, during the non-literate epoch, […] the gulf between poetic and prosaic could not subsist to the degree it does with us. The whole memory of a people was poetised, and this exercised a constant control over the ways in which they expressed themselves in casual speech" (134). Oral texts cannot be judged as isolated high-art objects that lie

outside of common experience. As Havelock explains, “Poetry was not ‘literature’ but a political and social necessity. It was not an art form, nor a creation of the private imagination, but an encyclopaedia maintained by co-operative effort on the part of the ‘best Greek polities’” (125).

When there is no way to store complex assertions or ideas, communication is needed. To communicate ideas in an oral culture is to keep them alive. This is why oral texts are such a key part of oral culture, as they are repositories of knowledge, experience and heritage. But in the case of new ideas, the bar of entry into a tradition of collective memory is set very high. New thoughts are harder to work through, as Ong argues:

> even with a listener to stimulate and ground your thought, the bits and pieces of your thought cannot be preserved in jotted notes. How could you ever call back to mind what you had so laboriously worked out? The only answer is: think memorable thoughts. In a primary oral culture, to solve effectively the problem of retaining and retrieving carefully articulated thought, you have to do your thinking in mnemonic patterns, shaped for ready oral recurrence. Your thought must come into being in heavily rhythmic, balanced patterns, in repetitions, or antitheses, in alliterations and assonances, in epithetic or other formulary expressions, in standard thematic settings (the assembly, the meal, the duel, the ‘hero’s helper’, and so on), in proverbs which are constantly heard by everyone and so that they come to mind readily and which themselves are patterned for retention and ready recall, or in other mnemonic form. Serious thought is intertwined with memory systems. (34)

The key phrase here is “think memorable thoughts” (34). As Ong explains, even if types of thought *could* be developed without these mnemonic structures there would be no benefit for oral speakers as they would disappear, both in their own minds and within the community. Thought must be placed into these formulaic patterns in order to exist. Proverbs and expressions such as “'Red in the morning, the sailor’s warning; red in the night, the sailor’s delight’ […] are not occasional [in oral culture]. They are incessant. They form the substance of thought itself. Thought in any extended form is impossible without them, for it consists in them” (35). This example leads to Ong’s key distinction between orality and literacy. The invention of the written word for Ong and his peers opens up this rigidity. However, this rigidity is a reflection of media-specific characteristics; literacy provides its own formal restrictions and affordances. The shift from print-based literacy to digital literacy also provides its own set of structures. Ong’s description that those in oral cultures had to “think memorable thoughts” (34) maps clearly onto what we would describe as search engine expertise. As established in chapter one, to formulate effective search engine queries is not a process of properly formulating a traditional question; rather, it is a process of iterating citations of language in which their answer might be written. To search effectively is to have a deep awareness of the language patterns of the source a user wishes to find. Those theorists, Ulmer, Kittler, and others, who argue that digital literacy is a fundamental shift generally describe such a process as one in which distinctions are broken down under the structure of programming. As Kittler articulates it:

> The general digitization of channels and information erases the difference among individual media. Sound and image, voice and text are reduced to

> surface effects, known to consumers as interface […] Inside the computers themselves everything becomes a number, quantity without image, sound or voice […] with numbers, nothing is impossible […] Instead of wiring people and technologies, absolute knowledge will run as an endless loop. (*Gramophone, Film, Typewriter* 1-2)

Kittler makes an important point about the tendency of the digital to absorb other media. However, in my view, this accentuates, rather than dulls, the characteristics of different media and nowhere is this clearer than when using Google's search engine. A Google search does not simply query an index of text specifically written for webpages. Google's index is filled with raw text and semantic metadata drawn from Google Books, Google Maps, Google Images, Google Shopping, YouTube and elsewhere. Formulating a query is a process of orientating a word or phrase towards a particular type of media. A version of Ong's "think memorable thoughts" (34) is in operation with query formulation when a user adopts the dialect to suit their frame of reference. This is not to say that all users are explicitly thinking through the framing of their queries, so that they follow how they might be phrased in a book rather than a blog. One of the lessons to be learned from Ong is that individuals in an oral culture need not be aware of the epistemological constraints that they are under in order to be affected by them.

To draw from Kittler's terminology, Google's search engine has opened up a huge range of differently mediated *discourse networks*, but their access still depends on a fluency of the cultural codes required for their understanding. Geoffrey Winthrop-Young defines Kittler's discourse networks by stating that "A culture is a large scale information machine, which, depending on the way the data inputs,

throughputs, and outputs are wired, produces basic notions as to why and to what end this machinery is supposed to function" (39). To describe Google as a technology *within* culture is to downplay the rootedness of 'culture' *within* a larger mechanism, a mechanism of human and non-human actors, cultural norms, and mediated expression. All of Google's activities, from its aims to scan every book, photograph every street, aggregate data of every dynamic pattern it can from influenza epidemics to traffic flow, all of these and more, present the case that Google's mission statement "to organize the world's information and make it universally accessible and useful" ("Google: About Us") is not meant lightly. George Dyson recalls, in *Turing's Cathedral*, his visit to Google where he interviews engineers about their projects. In particular, he discusses the aim of the Google Books project in scanning all printed media, to which an engineer expresses that "We are not scanning all those books to be read by people […] We are scanning them to be read by an AI" (312-313). Other interviews with Google staff also present the case that Google treats all of its disparate activities under the same banner of what Hillis et al. describe as "the search for a universal language, a universal method […] Leibniz's […] encyclopedia [sic] of *all* human knowledge" (100-101). In Steven Levy's study of Google, *In the Plex: How Google Thinks, Works, and Shapes Our Lives,* he documents a particular conversation with Sebastian Thrun, one of the founders of Google X, a research and development facility, about Google's self-driving cars:

> Critics charged that the project was a sign of Google's lack of focus – why was an Internet search company working on cars that drive themselves? Actually, the project was well within Google's wheelhouse. Since its earliest

> days, Brin and Page have been consistent in framing Google as an artificial intelligence company – one that gathers massive amounts of data and processes that information with learning algorithms to create a machinelike intelligence that augments the collective brain of humanity. Google's autonomous cars are information-collectors, scanning their environment with lasers and sensors, and augmenting their knowledge with Street View data. (Unlike human drivers, they always know what's around the corner.) "This is all information," says Thrun. "And it will make our physical world more accessible." (385)

The links between many of Google's projects become clear when they are placed into the context of gathering, analysing, and utilising information. This is also the case when tracing Google's acquisition and investment activities. For example, Alexander Halavais adds to these examples by highlighting "Google's investment in 23andme, a site that provides information about subscribers' DNA" (183-4) and their move

> into the real world, facilitated by RFID tags, GPS devices, and locative technologies leads us toward a "web of things," in which […] every object you own is locatable by a radio beacon, Google will be able to find your keys (under the sofa), your favorite novel (at your friend's house), and your spouse (at the pub). Do you want your search for kiwi yoghurt to be global, limited to your own refrigerator, or limited to any of the refrigerators in the building? The primary use of RFID-tagged objects will be to interact with the environment and with specialized systems, but given that search was about real-world

> objects way before it was about electronic documents on a worldwide network, we can expect that our physical environments will eventually become searchable. (184)

At the heart of Google's operations is a serious epistemological quest to model the world within a digital framework, much like the expanding imperialist map of Jorge Luis Borges' "On Exactitude in Science" "whose size was that of the Empire, and which coincided point for point with it" (325). The philosophical context in which Google operates is vast and it is essential to this study that its mission is taken seriously. Google's project is by no means the first practical attempt to map and harness the world's information – Paul Otlet and Henri La Fontaine's *Mundaneum* is a highly comparable twentieth-century analogue – but it has certainly become the most profitable. The economics behind Google's success will be addressed fully in chapter five; for the present discussion, it is important to note that although Google's project is a profit-driven capitalist venture its activities are still embedded within, and underpinned by, a long epistemological narrative. Therefore, I will now outline the epistemological positions taken by Plato and Aristotle, first, because they relate to the practical impact of literacy, and second, because their schemas underpin the two main conceptual perspectives that can be taken in regard to memory.

4.0 Ancient Modes of Memory

In the last chapter, Meno's paradox was outlined in the context of how subjects can form questions about subjects of which they are ignorant, in order to demonstrate the inherent paradox at the heart of enquiry. However, Plato's Socrates proposes a model of thought, anamnesis, in which Meno's statement is not paradoxical. Socrates outlines the basis of anamnesis in the following way:

> As the soul is immortal, has been born often, and has seen all things here and in the underworld, there is nothing which it has not learned; so it is in no way surprising that it can recollect the things it knew before, both about virtue and other things. As the whole of nature is akin, and the soul has learned everything, nothing prevents a man, after recalling one thing only—a process men call learning—discovering everything else for himself, if he is brave and does not tire of the search, for searching and learning are, as a whole, recollection. We must, therefore, not believe that debater's argument, for it would make us idle, and fainthearted men like to hear it, whereas my argument makes them energetic and keen on the search. ("Meno" 81d)

The idea of an immortal soul does not originate with Socrates. Many pre-Socratics, Pythagoras in particular, believed in similar versions of transmigration and this accounts for the matter of fact way in which Socrates introduces the immortality of the soul. Developing the idea that there is knowledge which transcends individual bodily experience and carries over from life to life, Plato argues for the idea of knowledge as finite as well as objectively structured. This, in turn, relates to Plato's

more general concept of ideal forms: the idea that every worldly experience is a mere shadow of a more fundamental abstract concept. Plato's theory of learning as recollection privileges the universal over the particular and downgrades the importance of sensual, particular, experience. This attitude is not unique to Plato; the notion of an unchanging world is a key part of pre-Socratic thought, particularly in the thought of Parmenides, for example. The Greek word for truth ἀλήθεια (*aletheia*) translates literally as "unforgetting" (Scanlan 30). Each life does not provide us with new knowledge but instead opportunities to rediscover the unconcealed or unremembered. However, although the word is used by early Greeks and literally translates the same, the meaning infused by oral thinking was very different from that of Plato's Socrates. As demonstrated earlier, Homeric thought embodies and can represent oral tradition and its unique thought structures. Christopher Long describes that

> In Homer, truth [ἀλήθεια] grows out of and remains ultimately concerned with the concrete lived experience of human being-with-one-another. From its beginnings, truth has always lived in and from interhuman being-together, even if already at an early age truth begins to stretch itself out toward the world of things […] truth itself does not originally appear as a relation between humans and things; rather, it emerges in dialogue between humans and it announces itself often in the form of a question […] thus for early Greeks, truth was at first neither a matter of the mind's relation to nature nor a property of mental representations, but an urgent question bound up with human being-together. (26-28)

Long argues that in Homeric poetry *aletheia* was particularly used in moments of conversation or questioning. In particular, it is used often in personal testimony when another character did not directly witness an event. We can then see how *aletheia* might have the same meaning in translation of “unforgetting” (Scanlan 30) to the Homeric Greeks and to the Plato. Truth in an oral culture, as demonstrated above, is established *between* persons. Phrases and proverbs stabilise ideas in the minds of people and between them through communication but they can never gain a truth outside of the people who have remembered such phrases. Therefore “unforgetting” to an oral culture refers to an action that, to a modern reader, appears rather literal and therefore less profound than Plato’s version; such an attitude demonstrates the influence of a literate bias.

Ong’s assertion that people in an oral culture had to “think memorable thoughts” (34), should highlight that the link between the memorable and the thinkable does not exist today in the same way. In a profound sense, the forgotten in oral culture is something that ceases to exist. That which cannot be remembered, then, is something that can never exist, or in other words, something that cannot be true. This is very different from Plato’s view that all knowledge exists at all times, just elsewhere (in the ideal realm). For Plato “unforgetting” represents the relationship between an individual (in their many lives) and a wider non-human existence. In Plato, the philosophical groundwork of memory shifts from one within community (or, in fact, that which *constitutes* community) to one that is individualistic and aims towards objectivity. Such a shift, argue Havelock and Ong, is due to Plato’s indebtedness to literacy. His version of *aletheia,* as well as the fundamental basis of his thought more generally, is due to his technologized mind: his interiority of the written word. It is this, as well as his critique of writing, that we turn to now.

4.1 Literate Plato: “Phaedrus”

One of Plato’s dialogues that recurs throughout the academic literature concerning technology is “Phaedrus”. The dialogue principally concerns rhetoric and oration, but it is the final section that philosophers of technology have most frequently addressed. It is in this section that Plato’s Socrates recites a story concerning the birth of writing in ancient Egypt. He tells of how Theuth, the god of writing, measuring and calculation, presents these and other arts to the king of the gods Thamus as gifts for the people of Egypt. Concerning writing Theuth states: “O King, here is something that, once learned, will make the Egyptians wiser and will improve their memory; I have discovered a potion for memory and for wisdom” (“Phaedrus” 274e). However, Thamus rejects such claims and replies that writing

> will introduce forgetfulness into the soul of those who learn it: they will not practice using their memory because they will put their trust in writing, which is external and depends on signs that belong to others, instead of trying to remember from the inside, completely on their own. You have not discovered a potion for remembering, but for reminding; you provide your students with the appearance of wisdom, not with its reality. Your invention will enable them to hear many things without being properly taught, and they will imagine that they have come to know much while for the most part they will know nothing. And they will be difficult to get along with, since they will merely appear to be wise instead of really being so. (275b)

Socrates continues to list the faults of writing, arguing that like paintings, written words “stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent” and “signify just that very same thing forever” (275e). Due to this repetition of the same, writing, when “faulted and attacked unfairly, [will] always needs its father’s support; alone, it can neither defend itself nor come to its own support” (275e).

In summary, Socrates’ criticisms are as follows. First, true knowledge can never be outside the mind; like a second-order version of the Cave metaphor, external knowledge will only be a shadow of a truer reality that exists in the mind. Therefore, the difference between written memory and memory within the mind is a difference of *kind* rather than of degree. Second, the illusion of real knowledge, created by writing, will cause ignorance. People will not only become lazy – a moral judgement is clear here – but also their minds will become less capable. Third, the combination of these two factors will produce individuals who are “difficult to get along with” (275b): bluffing know-it-alls. It is clear that Socrates does not believe in shortcuts to knowledge, in the same section of the dialogue he uses a metaphor of a farmer, planting his seeds at the wrong time of year and expecting a harvest in seven days. True knowledge is acquired slowly and those who have spent sufficient time developing their expertise should be respected. This links to the fourth aspect of Socrates’ critique. The written text cannot explain or rephrase itself. Compared with the communication between two interlocutors, this can lead to a text not functioning successfully, not truly getting the intended point across; in addition, a text can bluff: a clever-sounding statement can be written by a fool. This distinction between appearance and truth troubles Socrates and is due to Plato’s philosophy of ideal forms. Earthly knowledge is a shadow of an objective truth; the human mind mirrors

(imperfectly) the realm of forms. Socrates argues that the only true medium for knowledge is the “living, breathing discourse of the man who knows” (276a) as “It is a discourse that is written down, with knowledge, in the soul of the listener; it can defend itself, and it knows for whom it should speak and for whom it should remain silent” (276a). The centrality of the soul relies on the previous discussion regarding anamnesis, which suggests that although the external lives of people are temporary and change, the soul has seen everything already in its correct configuration. This is reinforced at the end of the dialogue in which Socrates prays aloud before he and Phaedrus return to the city: “O dear Pan and all the other gods of this place, grant that I may be beautiful inside. Let all my external possessions be in friendly harmony with what is within. May I consider the wise man rich. As for gold, let me have as much as a moderate man could bear and carry with him” (279c). The interior and exterior should reflect one another. The mind should be as beautiful as the external phenomena it represents and the external aspects of Socrates should be a reflection of who he really is internally. Considering wisdom as wealth uses external metaphors to describe internal realities and foregrounds the notion that wisdom should concern only that which is portable and internalised. Gold stored elsewhere cannot be used at the market, just as knowledge written on paper, stored at home, cannot be described as *your* knowledge.

4.2 Aristotle’s Sensory Approach

Aristotle’s theory of memory is different from Plato’s in two significant ways. First, Aristotle attends to the sensory world and, second, he distinguishes between memory and recollection. The first difference is a reaction to the kind of thinking

which underlies Plato's aviary metaphor: "The metaphor of the storehouse raises the question of how something can be found in the memory which has not entered through the doors of the senses" (Draaisma 28). Aristotle argues that, if we think of the mind as acquiring memories that can be recalled, then these must have developed out of our sensory experiences. Rather than seeing ideas as emerging into the mind through anamnesis,

> Abstractions and imprints of objects of perception must be part of the same continuum of perception and cognition. Aristotle asserts that unless one perceived things one would not learn or understand anything (DA III, 8 432a 3f). When one contemplates in the most abstract way one does so from an image. (Coleman 22)

Aristotle's teaching, that sensible phenomena were the source of higher concepts, became one of the founding principles of the Peripatetic school, fed into Scholasticism, and was rearticulated by Thomas Aquinas as the axiom that: "nothing is in the intellect that was not first in the senses" (2).

Instead of abstractions providing the model for imperfect copies, Aristotle's theory of memory posits an opposite approach. Individual sensory impressions are arranged, extrapolated and generalised in order to form abstract concepts: the individual images remembered are much more important to Aristotle than to Plato. This brings us on to Aristotle's conception of recollection. Recollection for Aristotle is not like natural involuntary memory it is, rather "a sort of reasoning process, described as a search, starting from one's thinking of something rather than from one's perceiving it. It is a deliberate undertaking. It involved a succession of

associated ideas [and] is a deliberate, self-motivated, autonomous process of search" (Coleman 23). For Aristotle, the act of recollection is creative. Rather than separating thought, memory, and creativity, recollection demonstrates how our memories are not stable and fixed: our ideas are not a selection of individual birds sat in an aviary. This idea prioritises individual agency and celebrates context. In describing recollection as active, Aristotle starts to detail the way in which students might better use their own memories. Aristotle describes the process of recollection as "starting in thought from a present incident, we follow the trail in order, beginning from something similar, or contrary, or closely connected" (Aristotle *De Memoria* 451b 18ff, qtd. in Carruthers, *Book of Memory* 79). However, this trail is not automatic and therefore forms the difference between memory and recollection; recollection is a process that can be improved, adapted, and trained. Aristotle's method of recollection was based on *place* (*apo topon*[5]). As Coleman describes:

> People are therefore thought to recollect, […] starting from places, and from such a starting point they associate to a succession of other things until they achieve the terminus of their search […] recollection is the process of finding the next or neighbouring item in a series it derives from the habit of *thinking* of 'things' in a certain order. (245 emphasis in original)

Such a description of active and personal recollection directly describes the kind of phenomenological report, outlined above, of the Google search strategy employed to find out the title and artist of a song. Although, as stated above, Google's overall project – "to organize the world's information and make it universally accessible and

[5] Literally: "from their sites" (see Lynch 74).

useful" ("Google: About Us") – deploys a certain Platonic logic, the process of *using* Google is much more in line with how Aristotle describes the process of recollection. The kind of memory that was useful in that example were the interrelated links between concepts. The more links a person can make between individual memories the more useful each memory becomes. For a user of Google's search engine, expertise relies on explicit search tactics as well as implicitly held expectations regarding the underlying structure of the web, as presented by Google. In addition, the increasing shift towards a mobile use of Google, on phones and tablets, intensifies the importance of context and place in memory. Users are more commonly engaging with Google in a range of environmental contexts; the image of a user sitting down at a desk in a hermetically sealed office space, or that cyberspace might exist in a realm divorced from the everyday, is becoming less realistic. However, at the heart of Google's project lies a tension between the philosophical views of knowledge, cognition, and memory proposed by Plato and Aristotle. Such a tension has existed within technosocial practices before; a notable example is that of the *Ars Memoria*, a technique for organising memory and knowledge, used for two thousand years.

The *Ars Memoria* is a set of memory methods or techniques that originally exemplified an Aristotelian conception of memory, although it predates him by some time. The narrative of the *Ars Memoria* is pertinent to the discussion of Google's philosophical underpinning for a number of reasons; one of these is that, although it initially embodied an Aristotelian attitude towards memory, the practice was taken over by those who wished to demonstrate that through rigorous application, such a system could exemplify Neo-Platonic objectivity. This narrative, that runs roughly from 400 BCE to 1600 CE (see Yates *The Art of Memory* 29-31), maps onto

Google's present situation, as described above, in which Google envisions that a complete enough aggregation of personal data (search strategies, book scans, street-view photography and other methods) could, through its AI, construct an objective collection of knowledge. In essence, Google's epistemological framework aims to transform the Aristotelian into the Neoplatonic; the history of the *Ars Memoria* demonstrates that this path has been trod before.

5.0 The *Ars Memoria*

The *Ars Memoria* represents a long tradition of memory training using certain techniques, methods, and frameworks. Our closest modern example of the *Ars Memoria* are the kinds of mnemotechnics, or memory techniques, used in contemporary memory competitions for recalling decks of cards, strings of binary digits and random historical dates. Today, displays of prodigious recall may appear disconnected from the practicalities of everyday life; the history of the *Ars Memoria* demonstrates how such an attitude is historically contingent and that attitudes towards memory and cognition develop in concert with technology. The techniques deployed by contemporary memory athletes represent variations on the set of methods outlined in three texts from antiquity: "*Rhetorica ad Herennium*"[6], "Cicero's *De oratore*", and "Quintilian's *Institutio Oratoria*" (Carruthers, *Book of Memory* 89-90). The main part of the *Ars Memoria*[7] is a skill of building, what are often termed, *memory palaces* and although there are many details for improving each set of memories the basic application is easily learned. The basic essentials of the technique consist in calling to mind a real place, often a building, and populating it with images that represent things to remember. The images are made to stick in the memory through a number of sub-techniques where the idea to be memorised is transformed through *personal* association into an image which is often bizarre, lewd, sexual, or comical. The *Ars Memoria* requires explicitly personal associations which are often unexplainable but which inexplicably stick in that individual's mind. The backdrop, or palace, must be mentally traversed in exactly the same order each time

[6] Although this text was accredited to Cicero for many years, modern scholarship has since moved away from this interpretation.
[7] Although repetitive, I continue to use the title *Ars Memoria* throughout this section rather than attempting to replace it for synonyms such as system, method, technique, technology etc. as the *Ars Memoria* is of interest *because* it cannot be adequately described as any of those.

for the links between the images to stay in the correct order. One of the main benefits of the *Ars Memoria* is that it can be used to memorise any order of magnitude. General ideas, like the previous example, can be used but also, and more commonly in its history exact words are memorised verbatim. In coordination with the precise principles of moving around the mental space in an exact order, this allowed practitioners to memorise whole books and even whole libraries. Even in situations where scholars had direct access to written texts, they would often memorise them using the *Ars Memoria* so that, in a pre-index or page number age, they could access any part of the library at will and quote from it verbatim. We have evidence of such activities, via Mary Carruthers, from the compositions of William of Ockham and John Wyclif. Carruthers details how even though both men were exiled and removed from their libraries, they continued to write as if their libraries were still available to them. While in exile in Munich, Ockham is quoted to have told a pupil of his, that "Complete knowledge about [the subjects of learning] should be patiently extracted and solidly built up" (qtd. in Carruthers, "How to Make a Composition: Memory-Craft in Antiquity and in the Middle Ages" 25). Indeed, Carruthers comments:

> Ockham did not educate himself with the idea that he might one day be exiled, nor as a student was he the captive of provincial schools, and, in consequence, deprived of ready access to libraries. His whole scholarly life until 1330 was spent in the greatest of European universities, his circle the most academic of the time. And still it is clear that he read to memorize and that in composing he drew extensively on the resources of his mental library. ("How to Make" 25)

In addition, Wyclif's works while in exile in Lutterworth "are filled with quotations from a variety of sources, too many to possibly be accounted for by the few books he had available" (25). These men were not exceptions, the use of the *Ars Memoria* within scholarly circles was widespread. As Frances Yates[8] describes: "Whilst it is important to recognise that the classical art is based on workable mnemotechnic principles it may be misleading to dismiss it with the label 'mnemotechnics'" (Yates, *The Art of Memory* 4) for it might make "this very mysterious subject seem simpler than it is" (4). As mentioned previously, memory in historical periods without mass printing or even access to resources such as paper for personal note taking takes on a moral role intensely related to one's humanity. The *Ars Memoria* then was not a set of tools or techniques but a way of structuring the mind in order to build a sense of an individual, which in turn allowed the creation of shared knowledge between people in order to establish a sense of community.

The *Ars Memoria* foregrounds a distinction between memory *of* things and associative links *between* things; one did not simply use it for rote recall, rather, it highlighted the order and relationships between groups of things, particularly those relationships important for a specific individual. Thus, the *Ars Memoria* enabled the memory of whole books, entire libraries, and even the perspective that a particular scholar took to these works. The texts held in the mind were not ordered within a system that aimed toward neutrality, such as an encyclopaedia or a library cataloguing system, but depended on the sensory life experiences of an individual tied-up with their personal reflections and dispositions. As Carruthers states: "The true force of memory lay in recollection or *memoria*, which was analysed as a variety

[8] This section draws heavily on Yates' *The Art of Memory* (1966) as her study represents the first and most detailed study of the *Ars Memoria* in modern scholarship.

of investigation, the investigation and recreation of knowledge […] To achieve this power, people educated themselves by building mental libraries" ("How to Make" 16). This reference to recollection refers to Aristotle's schema, in which recollection is highlighted as creative use of memory rather than rote recall. The internal workings of the method relied on outside structures: to practice the technique properly *real* buildings were needed. Hence, what seems like a personal system also became collaborative in the sense that it drew on individuals in their capacity as public citizens and caused the architects of physical buildings to become the architects of mental landscapes. Memory palaces as part of the *Ars Memoria* could not be shared. Many people memorised the same objects, which existed outside of each memory such as books, but the *Ars Memoria* only set forth a system of guidelines for the different ways in which such things could be memorised. Each individual use of the *Ars Memoria* would have been different: personal, emotional, practical and meaningful in a number of ways.

5.1 The *Ars Memoria*: Growth and Demise

After its use in Ancient Greece, the *Ars Memoria* was passed down into Roman culture and survived the sacking of Rome exclusively in the *Ad Herennium,* the other two texts becoming temporarily lost. Its techniques were then detailed and expanded upon by "Albertus Magnus and Thomas Aquinas [who] certainly knew no other source for the rules" (Yates, *The Art of Memory* 57) than the *Ad Herennium*. The influence of the *Ars Memoria* continued as it was adopted into the practice of the Dominican Order. This adoption was significant for the *Ars Memoria*'s later direction into mysticism, which derived – at least in part – from Magnus' influence within the

Dominican Order. However, the key figure for this study is Ramon Llull. Llull was a thirteenth century Neo-Platonist and a member of the Franciscan Order who spent much of his life on religious missions in North Africa to convert Muslims to Christianity. He is important to this study due to his creation of a system developed throughout his life but established most fully in *Ars generalis ultima* (*The Ultimate General Art*)[9] (1305). This *Art* was "a complex system, using semi-mechanical techniques combined with symbolic notation and combinatory diagrams" (Llull, 1) which would organise knowledge and solve questions and importantly would be applicable "to all fields of knowledge" (1). Borges' 1939 essay "The Total Library", draws on Theodor Wolff's suggestion that the idea of a total library "is a derivation from, or a parody of, Ramon Llull's thinking machine" (Borges, 214). Drawing a line of influence from the intellectual activities of antiquity towards Google's contemporary mission, Llull's ideas represent a historical convergence of the two discourses. Importantly, Llull takes the *Ars Memoria*, a profoundly Aristotelian method, and reinterprets it along Neo-Platonist lines. Llull's *Art* was not just a system of personal recollection, but a way of ordering everything in the universe and, in a deeply religious sense, organising an ultimate truth. Llull wanted to create a machine, which, knowing the truth of the universe, would convert Muslims to Christianity by providing unequivocal evidence for the truth of his faith. In doing so he wanted to turn disagreements between men of different languages and faiths into one of logical deduction.

In *Google and the Culture of Search*, Hillis et al. argue that Llull represents an important forerunner of Google's lofty and techno-deterministic ambitions. They describe Llull's *Art* as "an early attempt to do something now quite widespread –

[9] Henceforth referred to as Llull's *Art.*

substitute technology conceived as somehow free of human ideologies for the inherently messy and unpredictable ideological political sphere" (93). Hillis et al. also relate Llull's work to the original tale of the Tower of Babel, from which Borges' story draws its name. Llull's *Art* not only tries to reconfigure Christian arguments in logical terms but attempts to place all aspects of faith into one language, that of logic: "his stated goal is the production of 'truth', his aim that the *Ars* serve to advance the arrival of something akin to an information monoculture ('one language, one belief, one faith')" (93). The underpinning of this idea is that language creates cultural difference and in turn creates inaccurate interpretations of truth. The belief that linguistic variation simply represents a problem that can be overcome by programming is held by many at Google's headquarters, and is addressed in chapter four. Llull's ambitions were never realised as, in 1314, he was stoned by a group of Muslims in the city of Bougie and died the following year. His relationship to the longer history of the *Ars Memoria* is important, particularly in relation to how it developed and ultimately declined.

Yates, in her history of the *Ars Memoria*, argues that "Though the Art of Lull [sic][10] in one of its aspects can be called an art of memory, it must be strongly emphasised that there are the most radical differences between it and the classical art in almost every respect" (*The Art of Memory* 175). Nonetheless, Yates does consider it a defining step of the *Ars Memoria* as it transforms into the version known in the Renaissance. Three main differences exist between Llull's *Art* and the Classical *Ars Memoria*. First, as mentioned above Llull draws on a Platonic rather than Aristotelian structure, via Augustine, and as such "claims to know first causes" (175). Second, Llull's *Art* does not use images, as the *Ars Memoria* does and instead

[10] Yates uses this spelling throughout so [sic] applies to any of her spellings of Llull as Lull.

“Lull designates the concepts used in his art by a letter notation, which introduces an almost algebraic or scientifically abstract note into Lullism” (176). The images, as described earlier, being too focused on an individual’s own memories and emotions to be objective in the Platonic sense. Third, instead of a static memory palace filled with objects “Lull introduces movement into memory” (176). Llull used revolving concentric circles, which attempted to reconstruct the motions of the psyche and thought as dynamic and responsive. In this way, “The first art [*Ars Memoria*] is the more artistic, but the second [Llull’s *Art*] is the more scientific” (176). The mnemonic use of the principles of the *Ars Memoria* were deployed in the procedures of its use, memorising how to work the complex machine, rather than as a system which itself provided answers. The Llullian *Art* was “intended to construct a world of phantasms supposed to express approximately the realities of intelligible order of which our world is but a distant and imperfect copy” (Couliano 34). In doing so Llull’s *Art* presents an attempt at functional Platonism. The earlier classical *Ars Memoria* required that an individual constructed the memory palaces themselves. In this way, the things remembered would be situated within a deeper understanding developed by that practitioner. For images to stick they must be meaningful to that person; rhyming or word association could be used but only within a subjective context that made sense to a particular individual. Lewd and sexual images often worked exceptionally well, much to the distaste of the religious orders in which the *Ars Memoria* was used. But for this reason, many of the image prompts would stay private and relate deeply to that person’s own fixations, both conscious and unconscious. With the Llullian *Art,* the meaning behind how it worked was hidden. Someone, Llull himself for example, would answer questions using the machine linked with his own mnemonic patterning to produce an answer, which the observer

would accept as *the truth*. Therefore, the answers provided by the *Art* aimed towards objectivity while remaining mysterious.

The point at which this process of mysterious objectivity fully developed into mysticism lies with Giordano Bruno. Bruno draws on the classical *Ars Memoria* as well as the mechanics of Llull's system. Yates explains that

> Just as he [Bruno] converts the images of the classical art of memory into magical images of the stars to be used for reaching the celestial world, so the Lullian wheels are turned into 'practical Cabala', or conjuring for reaching the demons, or angels, beyond the stars […] Bruno's brilliant achievement in finding a way of combining the classical art of memory with Lullian thus rested on an extreme 'occultising' of both the classical art and of Lullism. (*The Art of Memory* 211)

Although Llull's version of the *Ars Memoria* had evolved from its classical uses it still bore some resemblance to its dialectical roots. Llull's *Art* answered questions, from practical to theological using the logic and deduction of mnemonic techniques. Bruno turned these elements of logic into ways to summon demons and contact the dead. However, as strange as it might be to a modern mind there were strong links between mysticism and science in the Renaissance, particularly among the followers of Hermeticism, of which Bruno was a practitioner. In addition to the often systematic, practical and experimental nature of Renaissance magic, the use of numbers also draws it close to the scientific method:

> Thus the Renaissance magic was turning towards number as a possible key to operations, and the subsequent history of man's achievements in applied science has shown that number is indeed a master-key, or one of the master-keys, to operations by which the forces of the cosmos are made to work in man's service. (Yates *Giordano Bruno and the Hermetic Tradition* 147)

In Yates' sense above, Bruno links to the scientific tradition as well as to the history of computing. Bruno's aim was to describe the world with perfect clarity but also to control it and use the underlying structures of number to tame and shape it. This is in part due to his Neo-Platonist perspective instead of an alignment to the classical Aristotelian roots of the *Ars Memoria*. In Borgesian terms, Bruno sets himself up as a librarian in the "Library of Babel" rather than an acolyte of "Funes the Memorious". In addition, Bruno was one of the pioneers in arguing for a heliocentric universe, which was part of the reason he was burned at the stake for heresy in 1600. This led to his legacy as a martyr of science standing up for truth against religious superstitions, a position that is defended in Hilary Gatti's *Giordano Bruno and Renaissance Science: Broken Lives and Organizational Power*. As Gatti argues, "Bruno remains so much a figure of the modern world [and] was among those who guaranteed the future of the newly emerging sciences" (19). However, an alternative account, exemplified by Yates' critical biography, suggests that Bruno was involved at a much deeper level with Hermeticism and it was this affiliation that the Catholic Church objected to. Rather than describing these two aspects as disconnected, Yates argues that Bruno's scientific method develops from his Hermetic beliefs, in particular his famous defence of a heliocentric universe:

> He patronises Copernicus for having understood his theory only as a mathematician, whereas he (Bruno) has seen its more profound religious and magical meanings […] Bruno's use of Copernicanism shows most strikingly how shifting and uncertain were the borders between genuine science and Hermeticism in the Renaissance. (*Giordano Bruno and the Hermetic Tradition* 155)

For Bruno, a scientific or mathematic understanding of the world developed out of the magical aspects of the universe. Far from using science as a tool for debunking mystical or magical experiences, as is common today, science acted as an extension of existing spiritual belief.

This belief of a Platonically structured universe figures into our current rhetoric of modern computing. The way in which current technologies are often discussed, Google included, perhaps represents a new era in which mysticism has crept back into (or, in fact, out of) the scientific tradition. John Durham Peters devotes a whole chapter of his *The Marvelous Clouds: Toward a Philosophy of Elemental Media* to "God and Google" and due to its resonance with the arguments of this chapter is useful to quote at length:

> Sergey Brin famously suggested that "the perfect search engine would be the mind of God." This half boast, half ambition puts Google into a long line of hieratic readers of the sky, and has a nice a touch of Kabbalah as well. It shows Google's membership in a distinguished family of religious media. Google's project is to build a temple to meet the sky, anchor remembrance, and serve as a canon of all knowledge. Its aim is nothing less than a

> metamedium that would be the guide for the perplexed of cyberspace. Google inherits the narrative of the priestly class that discerns the universe, renders order out of chaos, answers our entreaties, and invites us to take part in mantic acts of divination. From the unaccountably vast array of possibles Google provides the answer you seek, rather like fortunetelling and haruspicy or the priests who stood in the *templum* watching the sky for augurs and omens. Google is a clergy defined by its control over the means of inscription and retrieval—as clergies and priesthoods always have been. Google also picks up on the long romance that mathematicians have had with infinite and ultimate things. "The respective interpretation of the symbols 0 and 1 in the system of logic are *Nothing* and *Universe*," wrote George Boole. This was a variant of Leibniz's view of digital notation as shuttling between creation and the abyss—indeed, in the space where Google likes to shuttle. (333-334)

In addition to Google's grand ambitions, the company also employ a number of individuals who embody technospiritualism. The futurist Ray Kurzweil is currently the head of engineering at Google, as well as a figurehead in the transhumanist community, and a leading exponent of the idea of the Singularity. Kurzweil's thinking is as thoroughly technodeterministic as it is mystical; after the singularity, an event he sees as the fifth (out of six) epoch of history he argues will come a period where the "universe wakes up" (33). This is the point at which, as machines become spiritual, the universe becomes self-knowing, perfectly ordered and complete. In many ways, the period between Bruno and the present-day, rather than separated by a Kuhnian paradigm shift in regard to science and mysticism, might instead be considered as a brief hiatus, a momentary lapse of faith. It is this period that the

following section addresses, to complete this narrative of memory from orality to the present.

5.2 The End of the *Ars Memoria* and the Beginning of Modern Science

The decline of the *Ars Memoria,* from the sixteenth century onwards, is embedded within widespread social, cultural, and technological change. It was previously noted that some thinkers place the accountability of the decline of the *Ars Memoria* on the rise of print. However, it is unclear whether its downfall was a direct effect of print culture or, instead, an indication of a wider cultural shift. As Carruthers states:

> it is my contention that medieval culture was fundamentally memorial, to the same profound degree that modern culture in the West is documentary. This distinction certainly involves technologies – mnemotechnique and printing – but is not confined to them. For the valuing of *memoria* persisted long after book technology itself had changed. (*The Book of Memory* 9)

Other explanations for the downfall of the *Ars Memoria* focus more closely on the associated impacts of the Reformation and Counter-Reformation. As Ioan Couliano articulates, the *Ars Memoria* "worked so well that it is astonishing that it fell into disuse in the seventeenth century" (181). Well-practiced *Ars Memoria* would have been more useful than a well-stocked library. As the eighteenth-century Dutch poet Jan Luyken put it, "One book, printed in the Heart's own wax / Is worth a thousand in the stacks" (qtd. in Draaisma 38). Such an attitude forms the basis of contemporary popular critiques, by Carr and others, that describe externally situated, or stimulated,

memories as inferior to *natural* ones, although this chapter demonstrates that such a distinction is misguided. The *Ars Memoria* was the most powerful mode of memory, as well as being one of the most dominant, for at least two thousand years.[11] Furthermore, considering that a form of the *Ars Memoria* is used by contemporary mental athletes, these techniques have remained effective despite the wider social changes. For this reason, rather than describing its decline as a lack of usefulness, many critics have directly attributed its downfall to the iconoclastic activities of the Reformation and Counter-Reformation. As Couliano states, "The results of this iconoclasm are tremendous if we consider the controversies about the Art of Memory aroused by Bruno in England: ultimately, the Reformation leads to a total censorship of the imaginary, since phantasms are none other than idols conceived by the inner sense" (193). In this way, the disappearance of the *Ars Memoria* represents one of history's most extreme rejections of artificial memory. The lush ornament of the mnemonic techniques, combined with its increasingly occult uses, caused the techniques to be deemed irreligious. After the Reformation had passed, the state of Western intellectual pursuit came to be dominated by the development of modern science and, in particular, the philosophical framework of Francis Bacon. In many ways Bacon's "restructuring [of] the enterprise of scholarship" (A. Wright 131) left no room for the *Ars Memoria*; although it would have continued to be a powerful system (as shown by present-day uses) it existed in many ways as an archetype of the kind of thinking Bacon was trying to reject. This rejection is key to outline as, although it easy to understand that the Neo-Platonic mysticism of Llull and Bruno's respective arts would have died out, their demise is no reason in itself that the original classical version of the *Ars Memoria* could not have been readopted.

[11] This is based on the time between its first known description *Dialexeis* c. 400 BCE (see Yates *The Art of Memory* 29-31) and Bruno's death in 1600.

One of Bacon's most important texts *Novum Organon* (1620) was a direct attack on Aristotelian methods, its title being an explicit reference to Aristotle's *Organon*. In laying the foundations for what would become known as the scientific method, Bacon outlined four idols, or barriers to true understanding, which should be avoided: the idols of the cave, the idols of the tribe, the idols of the marketplace and the idols of the theatre. These are important in the context of this chapter, not only because they describe the underpinnings of later science and the development of experimental psychology, but also, they attack directly the kind of attitude that supports the *Ars Memoria*. I will briefly detail them, via Alex Wright's synopsis, in order to complete this history of the *Ars Memoria* and to act as a summary of ideas that provides the background to scientific memory research in the twentieth-century.

First, the idols of the cave, which Wright terms "The problem of subjectivity" (135), argues that we each process information with certain personal biases. This is certainly true of the *Ars Memoria*, as while it might be used by a number of different people to remember the same text, it is constructed using personal associations and therefore will, in Bacon's view, be coloured by the individual. Second, Bacon warns of the idols of the tribe, that of being human and the sensory constraints that entail. Using "logical reasoning we can begin to compensate for our sensory limitations" (135). This again goes firmly against the Aristotelian model of the *Ars Memoria*, where the sense of sight is privileged over other modes of understanding. Third, Bacon outlines the idols of the marketplace, which addresses language: "Words are imperfect vessels, pale approximations of direct experience" (135). Whilst in the *Ars Memoria* images are selectively memorised, these often represent words directly (as mentioned above), and therefore Bacon's attack still applies. There is no way of storing a graph, chart or illustration in one's mind; only the accuracy of the printing

press can directly reproduce mapped quantitative difference. This also prefigures a substantial aspect to later practical studies of memory where memory is treated numerically, a far cry from a classical understanding of memory. Finally, Bacon outlines the idols of the theatre, the "problem of belief [in such things as] Mythologies, religious stories, or ideological convictions" (135). This final idol certainly excludes Llull and Bruno's attitudes, but neatly points towards Google's self-clarification as an organiser of knowledge, free of ideology.

The other important aspect of the *Novum Organon,* for our current concerns of memory, lies in his restructuring the faculties of the mind. Bacon made

> memory, reason and imagination the basis of his scheme, allocating history to the category 'memory', […] philosophy to 'reason', and poetry to 'imagination'. An examination of the curriculum, the library and the encyclopaedia in the seventeenth and eighteenth centuries suggests that Bacon's reclassification was the most successful of the various attempts made at this time. (Burke 99)

Bacon's scheme came to dominate through its self-reinforcing network of technosocial institutions of the curriculum, the library and encyclopaedia. The Dewey Decimal System, for example, follows Bacon's scheme and still represents one of the main contemporary top-down organising systems that Google's bottom-up approach challenges.

As John Scanlan outlines, the "hold of the Baconian view of memory and history began to loosen with the rise of historicism (the professionalization of history as a discipline) in the nineteenth century, but still held enough sway that it could be target of attack for […] R. G. Collingwood" (32) in the 1930s. There are significant

elements of Baconism still with us today. Although history as a discipline has moved away from Bacon's model, it has done so in a manner that also moves further away from the classical conception, outlined above. The professionalisation of history, in Collingwood's view, prioritises external evidence. Documents and artefacts are better than personal memory and better still are disembodied data. As Bill Schwartz argues, this conception of history is linked to a modern understanding of the past: "The intellectual practice of history, in its emergent forms was in part devised as a counter to the wayward, indeterminate workings of modern memory […] In this scheme of things, subjective time, the time of the everyday and of the self, memory included, could appear *only* as dysfunctional" (43). This increasing prioritisation of the impersonal created ripples throughout the early efforts of Psychology to measure cognition and memory, from the work of Robert Hooke (1635-1703), an original fellow of the Royal Society and their curator of experiments, to Hermann Ebbinghaus (1850-1909) and beyond. These modes of memory certainly influence the initial studies regarding transactive memory. Although the studies treat memory in various nuanced ways, as shared and situational for example, the methodologies of Sparrow et al. focus on the recall of trivia and various impersonal facts. That these results can then be measured, and tallied numerically for comparison, speaks to this historical lineage.

Hooke placed an emphasis on attention. He assumed that people's capacity for forming memories was generally the same; it is how much effort is put in to remembering ideas that is the key to the creation of a clear memory. This attitude opened up the possibility of studying the capacities and limits of human memory, as a generalizable category. Hooke was also a materialist, which creates another great difference between his theories and classical understanding. As Draaisma explains:

> Hooke saw the ideas stored in the brain as truly material entities, and in so doing, introduced a new type of question into theories of memory. For what is the *rate* at which these ideas are formed? What is the *number* of ideas in the memory? And what is their *location*? In traditional spiritual theories, memory was interpreted as a non-spatial entity and such questions hardly made sense. St Augustine took memory to be a *quasi*-space, "an inner place – though it is wrong to speak of it as a place". The fact that every human being gathers innumerable quantities of memories in his lifetime will not lead to a lack of space, since memory has no physical limits: "it is a vast, immeasurable sanctuary. Who can plumb its depths?" (57)

Hooke began to ask questions of mental storage, many of which resonate with contemporary digital metaphors for cognition; he concluded that attributing memories a size and location meant that the brain must have a limited capacity. Furthermore, if memory is reliant on attention, then individuals can only make so many memories per day, and by extrapolation, per lifetime. This kind of thinking maintains that memory is quantifiable and, even before computing, can be used to describe anything in terms of its size in the mind. Draaisma notes that Hooke is perhaps the first thinker in a tradition of attempts to calculate the capacity of the mind in quantifiable terms (after a number of calculations Hooke settles on an estimate that by one hundred years of age an individual might make four million memories). In reference to Hooke's calculations, Draaisma notes that

> anyone tempted to make sarcastic observations on the laborious calculations regarding the number of stored ideas in the warehouse should look at the way in which the number of 'bits' in the human brain is currently calculated. The author of an article giving an overview of the field, Landauer[12], uses a method essentially identical to Hooke's […] Landauer arrives at an 'input rate to long term memory' of 1.2 bits per second, where Hooke's estimate had been one mental image per second. For someone aged seventy that gives a total of 1.8 billion bits. (61)

This is not to argue that all cognitive science and psychology consists of a footnote to Hooke's thinking. However, Hooke's work acts as a catalyst for a certain narrative, which uses the computer as a model or metaphor for memory. The figure here who is the theoretical descendent of Hooke, and the father of contemporary experimental psychology of memory is Hermann Ebbinghaus (1850 – 1909). Ebbinghaus was not the first to draw conclusions about memory through experiments but "Ebbinghaus does have the honour of being the first to design and conduct an experimental *programme*" (93). Ebbinghaus invented a "simple technique that in various forms has been a stable psychologist's tool ever since – that of the nonsense syllable" (Steven Rose, "Memories Are Made of This" 199). Ebbinghaus used himself as his experimental subject and proceeded to test how well he could learn and recall these nonsense syllables such as "HUZ; LAQ; DOK; VER; JIX" (199, capitalisation in the original study). Throughout his research programme, Ebbinghaus built the foundations for the kind of research of the future, as well as provide much of the vocabulary now commonly used, both scientifically and colloquially, such as the

[12] See Thomas Landauer's "How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-Term Memory".

distinction between short and long-term memory. An important point to stress is that we might describe Ebbinghaus as a proto-behaviourist, as, “what mattered in experimental research into memory were measurements and numbers, not the links between memories and neurons” (Draaisma 98). Many other nineteenth-century scientists were focused on the anatomy of the brain, for example, locating where the centres for speech or emotion might reside. Ebbinghaus did not wish to explain the inner mechanics of the brain, but rather understand how it behaved. This kind of framework judges whether information has been successfully recalled, and to what degree of accuracy. The model of short and long-term memory develops throughout the history of psychology to become the Atkinson–Shiffrin memory model (1968) or the modal model as it is often called.

In outlining the differences between short and long-term memory Steven Rose refers to the influence of the information or computer model for cognitive science.

> The temporal distinction between short-term and long-term memory, the evidence that short-term memory is labile and easily disrupted, whereas long-term memory seems relatively protected, suggested that it must depend on some structural remodelling of the patterns of neural connections within the brain, engraving memory in the brain in a manner analogous to that of inscribing a magnetic trace on a tape or a CD that can subsequently be replayed, invoking the original material. The seductive metaphorical power of computer “memory” has been influential in shaping thought on this question. (202)

The stability of memory is rather misleading, as was outlined at the start of this chapter in regard to the constructivist approach to recall. As Rose goes on to argue, long-term memory has proven in a number of experiments to lack the stability that it is often credited with. In a review of the history of psychology of cognition Sutton et al. describe a number of critiques of the established modal model, stating that

> Despite the widespread adoption of this basic computational view of memory, recent theory and research recognize that memory is more complicated than this model might suggest, especially memory of personal experiences or emotional material […] memory has broader functions, such as maintaining our sense of self, regulating emotion, motivating and directing future action, and helping us to promote and maintain relationships with others. (213)

Such an attitude draws us back around to the psychological studies with which this chapter opened, in which the methodologies for studying transactive memory prioritised the memorisation of discrete facts that could be stored either in folders on a computer or in the mind. This connection should not be viewed as a dismissal of the work of Sparrow et al. and other transactive memory researchers, instead it should illustrate that technological metaphors are intertwined at all levels with attempts to describe the mind and its functions. It also foregrounds the way in which available methodologies shape the kinds of behaviour that can be studied. The above quotation by Sutton et al. regarding the broader functions of memory should fit with the description of memory developed from the examples throughout this chapter: memory in oral culture, the use of the *Ars Memoria*, and the phenomenological description of dynamic recall using a search engine.

Conclusion

The debate over whether Google could be considered a source of transactive memory is based on behaviourist observation; such experimentally-evidenced psychological insights do not provide much comprehension regarding the nature of *how* memory functions. Within a dominant discourse of computational models of memory and information, appeals from Sutton et al., to explore the affective and subjective aspects of memory are difficult, both in terms of how the wider discourse is framed and the methodological tools available for such research. The conclusions of the opening studies demonstrate that there is much to explore regarding a contemporary technologised memory. That individuals have a tendency to forget information that they think they will have access to later on, even if they actively try to remember it, has important consequences for Google's role in our current context. The historical narrative of this chapter shows that memory has always functioned in concert with technology; therefore, we should be sceptical of alarmist claims which assert that the contemporary situation presents us with unprecedented challenges. However, this does not mean that the kinds of studies addressed at the opening of this chapter, regarding transactive memory, should be disregarded. Such reports, when considered alongside a historical awareness, can enable a much more nuanced and specific discussion of search engines and their effects.

This chapter highlights that memory has always been described and defined through technological metaphor, whether that of a wax tablet, an aviary, a codex, a photograph, magnetic tape, a CD, or a hard drive. It is clear that conceptualisation of something so abstract as memory requires some allegorically theoretical basis and scholarship of new technology needs to acknowledge these underlying attitudes. In

addition, the management executives, engineers, and marketing teams of technology companies are influenced by, and continue to perpetuate, similarly inherited conceptual frameworks. Although Google often presents itself as free from ideology, its attitudes towards, and effects on, memory are situated in a long philosophical tradition. Memory is always embedded in social context and enacted through technological interaction. This means that Google's responsibilities are not new, although the scope of how many individuals are impacted by their perspective demonstrates a marked shift in power and control compared to previous technological moments. As shown in chapter one, Google designed their methods, such as their PageRank algorithm, to provide a model of the existing structures of influence on the web. However, this model should not be overdetermined as representing Google in a neutral role. Due to its global influence, Google have the power to dictate the kind of behaviours that search engines enable; should this ubiquitous technology be optimised for recall or discovery, should it direct users towards ideas and issues that fit neatly into their existing epistemological landscapes, or should they provide heterogeneous perspectives? These vital questions will be further explored throughout the following chapters. Chapter three demonstrates how Google's Autocomplete tool functions as a kind of collective memory, storing attitudes and outlooks for particular individuals in various contexts. In the way that the technosocial practices of memory discussed in this chapter relied on the dynamic use by a particular community, the following chapter shows how Google's organisational methods depend heavily on how its users interact with its search engine. In particular, chapter three draws attention to the automated methods of aggregating these collective memories and online actions. This chapter has demonstrated the uneven distribution of power over the methods and content of

memorialisation, from oral singers to the impact of the Reformation on illustrative memory, and the rise of modern science. However, as a private company, Google's level of authority and its power over individuals is unique. Chapter four returns to this issue and in particular draws on the distinctions between Platonic and Aristotelian notions of thought and memory, outlined here, in order to better contextualise contemporary debates regarding Google's role within the wider digital ecology. Finally, the economic incentives that emerge from Google's influence are most extensively developed in chapter five. In summary, this chapter establishes the historical and philosophical narratives that underlie the discussions of the following chapters. It establishes that the interaction between the technological and social that search engines embody has a historical precedent, while emphasising the specificity of those characteristics that are unique to contemporary digital culture.

Chapter Three:

Autocomplete

Figure 1. An example of an Autocomplete suggestion completing the stem [Google and I are so close that we finish each other's sen]. Search performed 1/08/2013.

Introduction

This chapter engages with Google's tool known originally as Google Suggest, later as Google Autosuggest, and finally as Google Autocomplete.[1] The tool comprises a drop-down menu of suggestions from which a user can select to either complete or replace their initial query stem (see fig. 1-3).

[1] For simplicity, the designation Autocomplete will be used throughout, even though this phrasing is anachronistic at times.

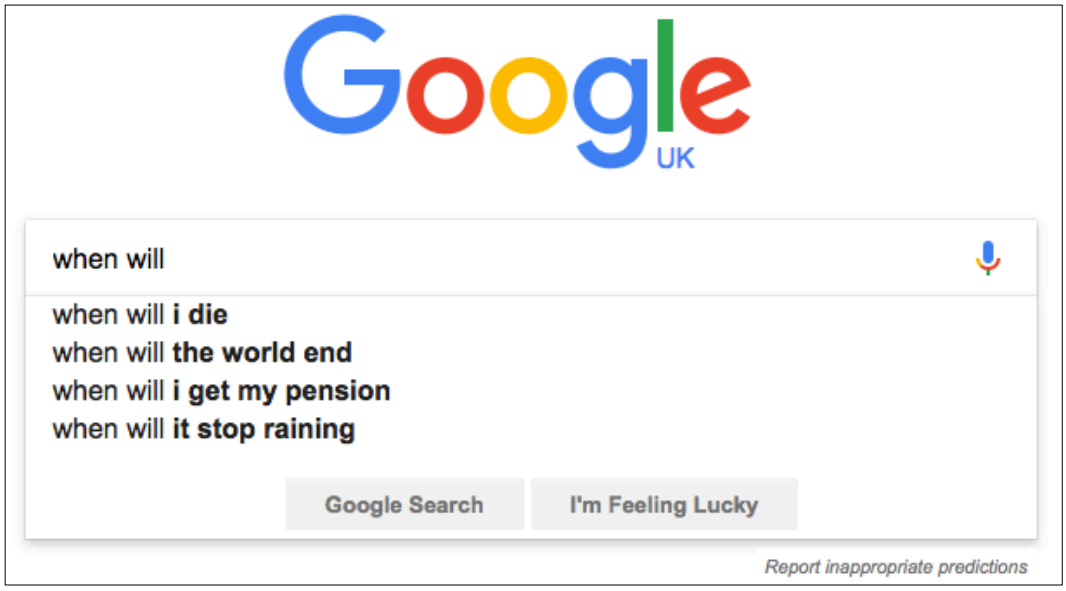


Figure 2. An example of Autocomplete suggestions for the stem [when will]. Search performed 9/8/2017.

Figure 3. An example of Autocomplete suggestion that rephrases a stem to more accurately reflect grammatical word order. Suggesting a replacement of [Christmas how long u] for [how long until christmas]. Search performed 9/8/2017.

Google Autocomplete is a compulsory tool that activates whenever a user begins their query; all other search engines such as Bing, Yahoo!, Baidu, Yandex and even meta-search engines such as Dogpile, have copied the functionality of Autocomplete in their search engine interfaces. Due to its central role in all aspects of web search, it represents an important part of understanding search engines. It has also become a staple for all web-based search (search bars on the websites of Facebook, YouTube, Wikipedia, LinkedIn and Twitter all provide input-text suggestion tools, visually similar to Autocomplete's dropdown box). Autocomplete has entered a realm of ubiquity throughout various digital contexts to the extent that its logic can be described as a "microinteraction", a term coined by software designer Dan Saffer. Liz Gannes, in a history of Autocomplete, describes how once design decisions become microinteractions "You wonder why things that don't have it are broken [and] only notice it when it messes up". This chapter specifically focuses on Google's Autocomplete, however, it is essential to acknowledge its broader legacy in contributing to widespread expectations throughout various areas of human-computer interaction.

I argue that Autocomplete is the aspect of Google Search that attracts the largest public interest; however, it has received minimal academic attention, in comparison to other aspects of search engines. Google provides very little information regarding how the tool functions, which has led to a number of misconceptions to be established and perpetuated through various public and academic discourses. These misconceptions are important not only because Autocomplete is a widely used tool on the web, but also because judging Google's liability for Autocomplete's suggestions should be tied to a specific technical understanding. The consequences of Google's potential liability have wide-ranging

implications for many areas of the digital environment. As tools for automating and aggregating content become more complex, it is essential that the various kinds of discourses attending to them – legal, political, social, cultural and economic – are based upon an accurate understanding of their operation. This chapter's original contribution lies in challenging these misconceptions, highlighting the wider issues that are at stake, and collecting together the available evidence regarding how Autocomplete functions, in order to make an argument concerning the importance of algorithmic logic and increasing centrality of neural network machine learning.

The structure of this chapter is as follows. First, I substantiate the claim regarding Autocomplete's large public interest and minor academic attention by drawing on engagements with Autocomplete in popular culture and providing an academic literature review. Second, the history of the tool is outlined while also summarising the insights provided by Google into how Autocomplete operates; this is elaborated through a survey of traditional probability programming, the relationship to Google's wider neural network machine learning initiatives, and first-hand examples that demonstrate that the basis for Autocomplete's suggestions is deeply misunderstood in both academic and public discourse. Third, this chapter details the controversy surrounding particular Autocomplete suggestions that perpetuate various negative stereotypes. By drawing on examples and the conclusions from the previous section, I argue that Google's neural network machine learning programme is the cause of these suggestions and represents one part of a company-wide discourse, regarding the shift from algorithms to machine learning. Such a claim regarding Google's current trajectory has been made by journalists but has not been addressed within academia. This discourse presents an opportunity to discuss the liability of search engines, highlighting the technical differences between algorithms

and machine learning, and to ask whether or not Google have a social responsibility, in regard to the establishment, consolidation and repetition of stereotypes or wider derogatory perspectives. Fourth, this chapter analyses the importance that speed plays in Autocomplete, particularly in regard to notions of play and critique within contemporary capitalism. Finally, this chapter describes Google's criterion of *relevance* in the context of Autocomplete as well as its connection to personalisation and other relative contexts, such as location. Chapter four focuses directly on Google's metric of relevance in greater depth, outlining its place within a wider philosophical tradition and its use as a quantitative benchmark in the discipline of information retrieval; therefore, this chapter sets out an enquiry and establishes a set of questions that will continue to be explored in chapter four.

1.0 The Public Fascination with Autocomplete Suggestions

A key aspect of this chapter is that, almost unanimously, the public, the media, and academics in various fields have a false understanding of what Autocomplete suggestions represent; it is erroneously repeated over and over that Autocomplete suggestions are a reflection of the most searched queries. Such a claim has been enhanced by the lack of information released by Google about the tool. However, Google have never stated that the suggestions are based solely on the metric of search query volume. Search volume is part of the Autocomplete's algorithmic criteria, but I provide evidence throughout this chapter that it is not the main factor and, in some instances, may not feature at all in the suggestions. However, due to the widespread misunderstanding that Google Autocomplete provides a direct reflection of query volume, and consequently the questions and ideas web users are most interested in, it has gained a status as a kind of digital oracle that supposedly knows us better than we know ourselves. This is particularly the case in popular culture. Online games such as Google Feud and idiots.win (see figs. 4-6) replicate the structure of the television game show Family Feud in which players guess the most commonly answered responses to a question.

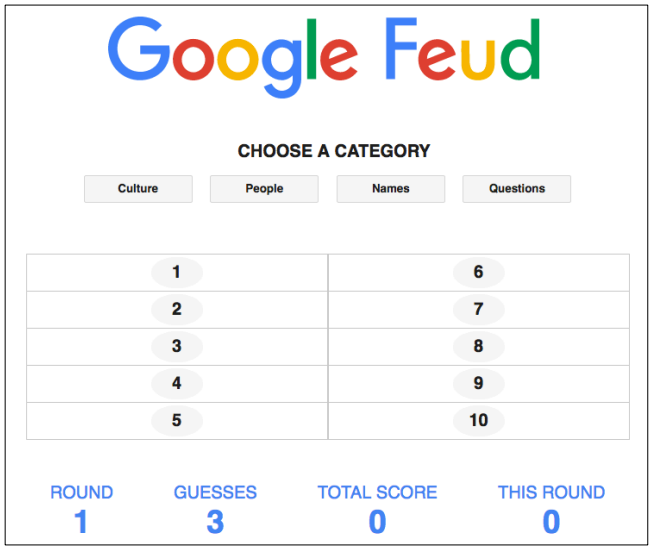


Figure 4. Screenshot of the game Google Feud, designed by Justin Hook, accessed 10/9/2017.

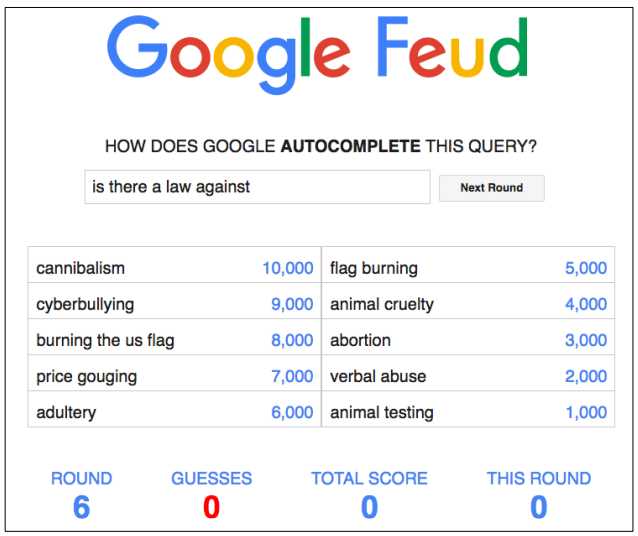


Figure 5. A screenshot of Google Feud mid-game, in which the player is asked to guess the correct Autocomplete suggestions for the query [is there a law against], designed by Justin Hook, accessed 10/9/2017.

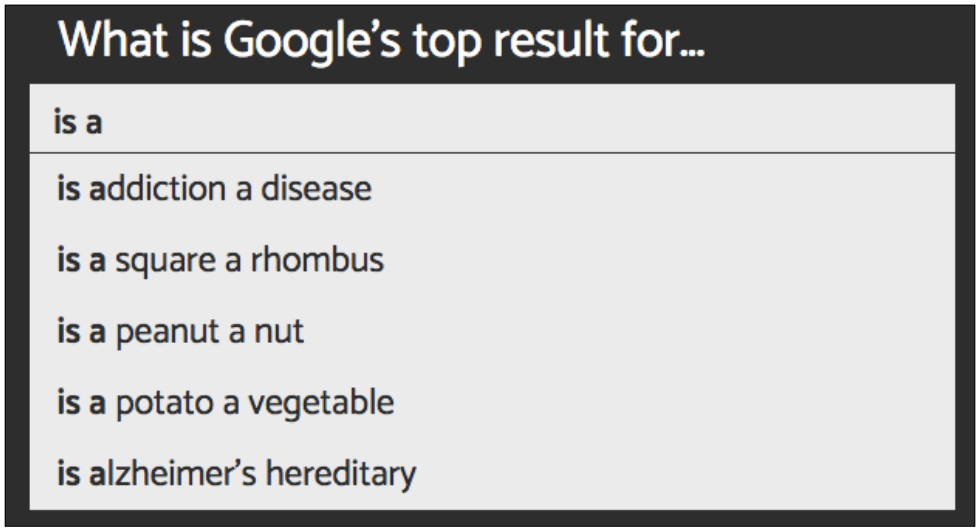


Figure 6. A screenshot of the game idiots.win, in which the player is asked to complete the Autocomplete suggestions for the query [is a] designed by "Lemon", accessed 10/8/2017.

The attitude that Autocomplete suggestions are a representation of public opinion, in light of the fact that many top suggestions are strange or unusual, has sparked great interest. An example of this can be seen in a number of genres of YouTube videos; one genre, "Autocomplete Interviews" established by *Wired* magazine, consists of interviews with various celebrities, in which the questions asked are not devised by a human interviewer, instead, they are derived from Autocomplete suggestions associated with their name. The misunderstanding that these suggestions are a direct indication of search volume is perpetuated through their titles, such as "Matt Damon & Julianne Moore Answer the Web's Most Searched Questions" (see figs. 7 and 8).

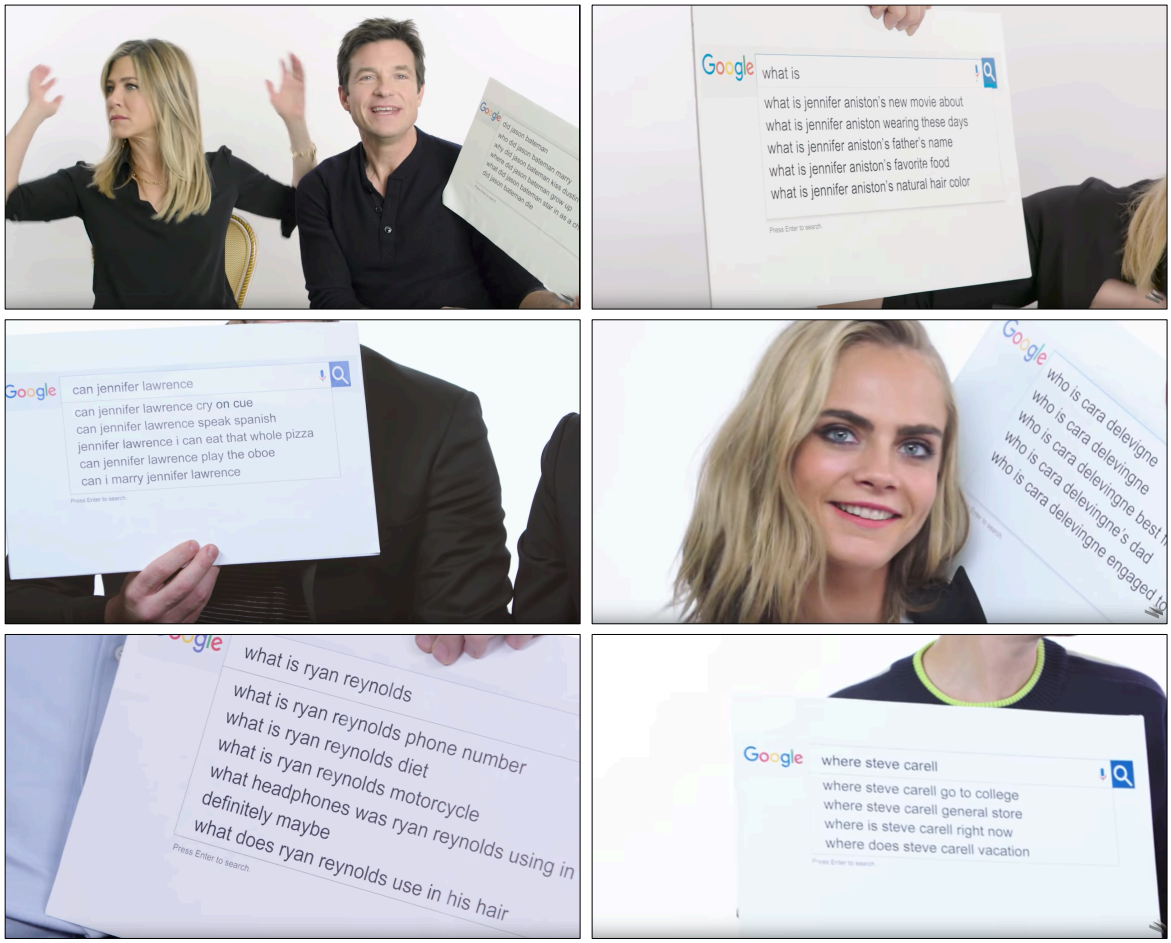


Figure 7. Stills from *Wired*'s "Autocomplete Interviews" accessed 11/8/2017.

As of 10/8/17 *Wired* magazine's series of thirty videos had cumulatively garnered 80,504,433 views on YouTube. However, such interest in Autocomplete queries is not simply related to celebrities; another genre of Autocomplete videos consists of vloggers simply reading out and reacting to queries suggested by Autocomplete (see fig. 9), from their standard searching or through playing Google Feud or idiots.win. The thirty most viewed English language videos of this type collectively had 84,051,052 views on YouTube, as of 10/8/17. The videos reflect the often strange nature of Autocomplete suggestions, with titles that recur around the question, "what is wrong with humanity?".[2] This fascination develops from the notion that

[2] This question is the title of a video by "DanAndPhilGAMES" listed in fig. 9.

Autocomplete reflects the kinds of questions people ask Google. Although, as will be demonstrated later in this chapter, this direct link between queries and suggestions is unfounded, Autocomplete suggestions *are* an insight into the way that Google collects and organises language. In this way, studying Autocomplete might not provide information on current search users, but it can illuminate a number of Google's activities, from its Google Books scanning project to its increasing reliance on machine learning neural networks.

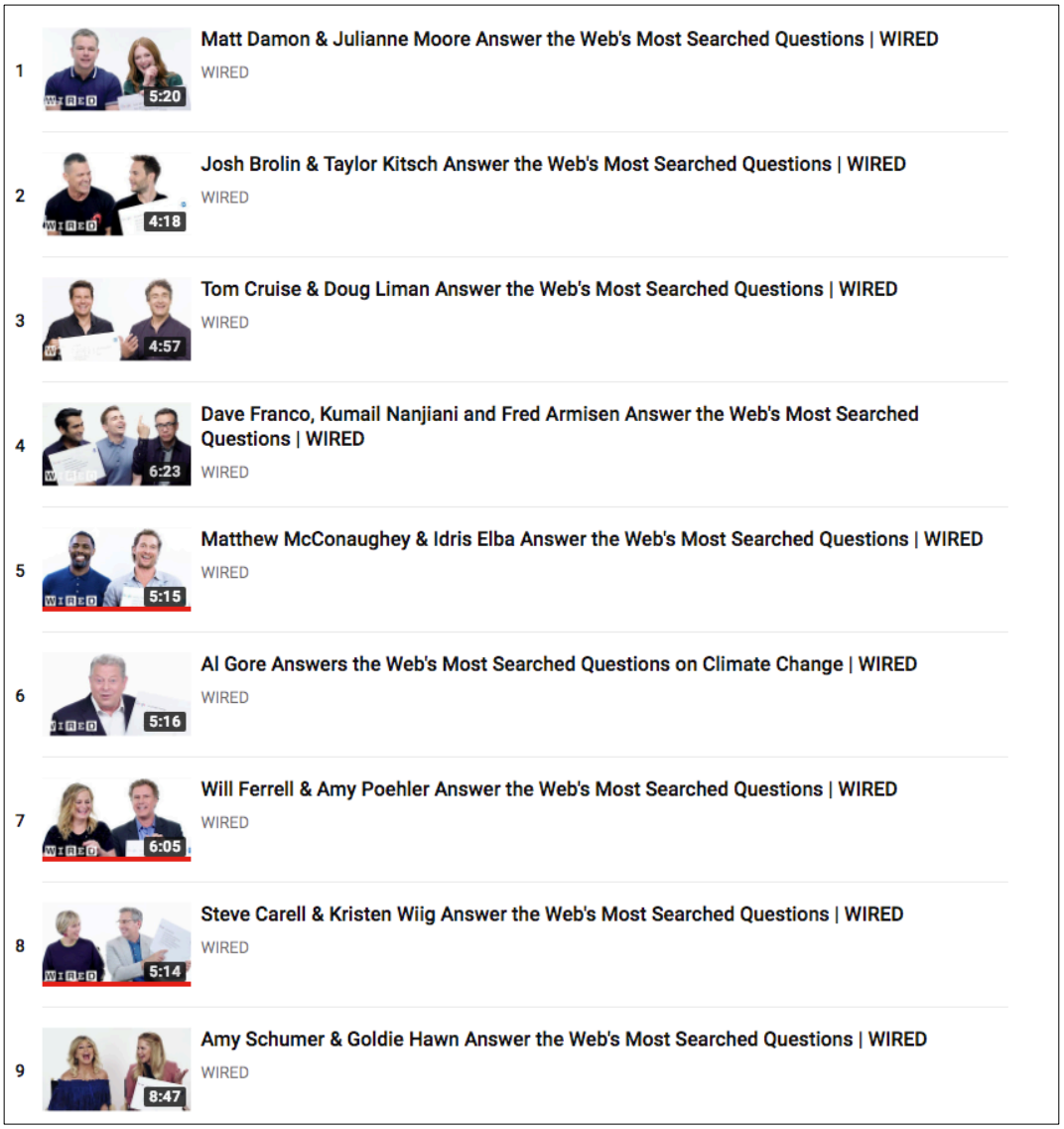


Figure 8. Examples of *Wired*'s titles for their "Autocomplete Interviews" Accessed 14/11/2017.

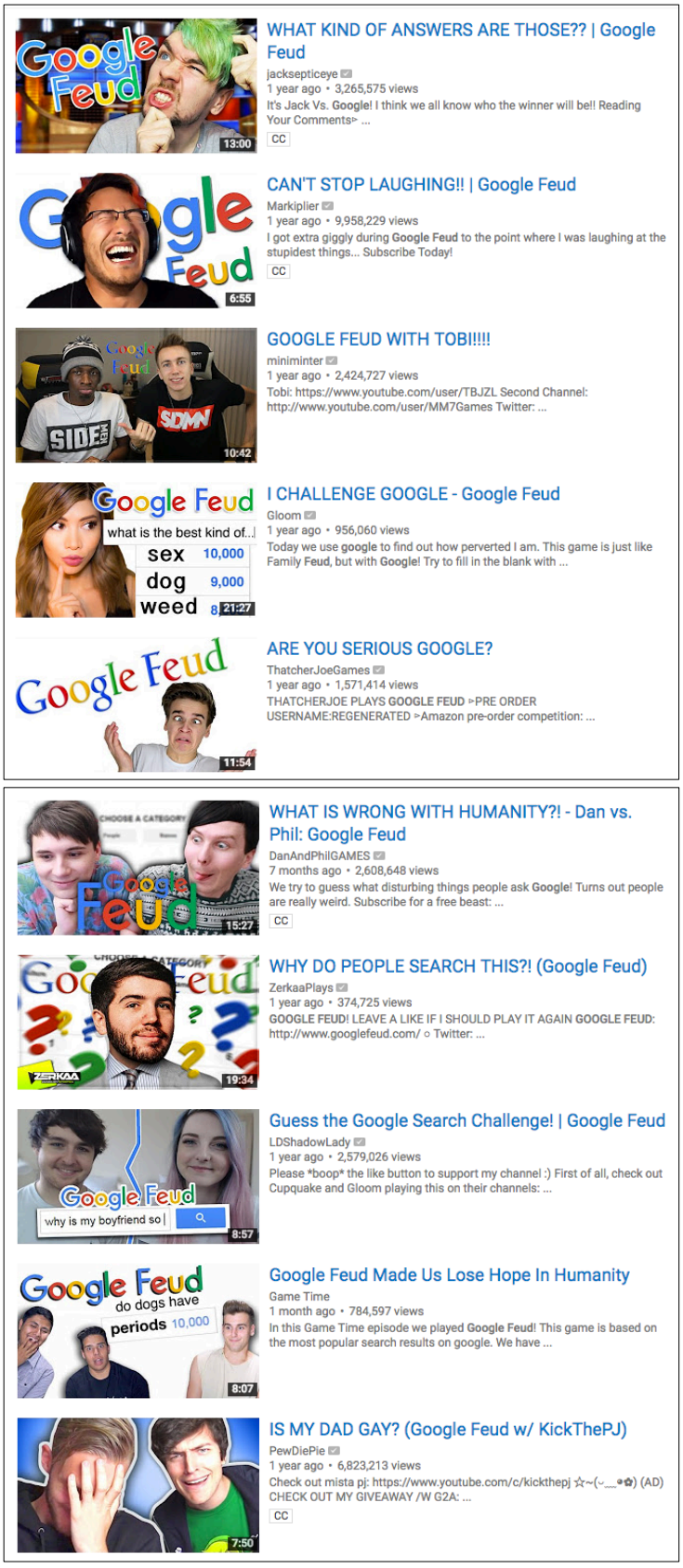


Figure 9. A selection of YouTube videos that consist of reactions to Google Autocomplete, Google Feud, and idiots.win Accessed 11/8/2017.

1.1 Autocomplete's Lack of Academic Attention

Autocomplete has become a public fascination on a scale of its own; there is not another search engine feature that receives anything like as comparable attention. However, academic enquiry is almost non-existent. Autocomplete has been available as an experimental feature since 2004 and as a compulsory component of Google Search since 2008; the following key, journal special issues, edited collections, and monographs make no mention of Autocomplete: in 2007, *The Social, Political, Economic, and Cultural Dimensions of Search Engines* (Hargittai ed.); in 2008, *Die Macht der Suchmaschinen / The Power of Search Engines* (Machill and Beiler eds.), *Web Search: Multidisciplinary Perspectives* (Spink and Zimmer eds.); in 2009, *Search Engine Society* (Halavais), *Deep Search: The Politics of Search Beyond Google* (Becker and Stalder eds.); in 2011, *In the Plex: How Google Thinks, Works, and Shapes Our Lives* (Levy); in 2012, *The Googlization of Everything (And Why We Should Worry)* (Vaidhyanathan), *New Perspectives on Web Search Engine Research* (Lewandowski ed.); and in 2013, *The Dark Side of Google* (Ippolita). In addition, the engagement with Autocomplete in journal articles is minimal. "'Why Do White People Have Thin Lips?' Google and the Perpetuation of Stereotypes Via Auto-Complete Search Forms" (Baker and Potts) addresses the kinds of stereotypes reinforced in Autocomplete but lacks any engagement with how the tool operates. In addition, there are journal articles from the discipline of law that address Autocomplete, however, all of them address whether or not Google is legally liable for the Autocompleted queries.[3] The only other texts that discuss Autocomplete are

[3] For examples, see "Google's Autocomplete Function – Is Google a Publisher or Mere Technical Distributor? German Federal Supreme Court, Judgment of May 14, 2012 – Case No. Vi Zr 269/12 – Google Autocomplete" (Peifer), "Search Engine Liability for Autocomplete Suggestions: Personality,

Google and the Culture of Search (Hillis et al.) and three chapters in *Society of the Query Reader: Reflections on Web Search* (König and Rasch eds.). These texts address the theoretical effects of Autocomplete but lack an in-depth engagement that tries to relate technical function, usage, and social impact. Kylie Jarrett, in René König and Miriam Rasch's 2014 collection, offhandedly refers to Autocomplete as part of Google's search interface that reinforces the "lie [that] is given to the concept of search as a transparent mediator of information on the web" (23). As will be shown throughout this chapter, although Autocomplete sometimes reinforces an illusion of neutrality, many of the suggestions achieve the opposite result by providing heavily subjective suggestions that in the case of the previously mentioned YouTube videos can cause bafflement and amusement. Anna Jobin and Olivier Glassey, in the same volume, address Autocomplete in the following way: "Google's autocompletion will suggest words and expressions to us before we finish typing. Thus, these algorithms mediate semantically between what we mean and which words we will use to describe it. All of them are so-called 'linguistic prosthesis'" (158). The notion of linguistic prosthesis is potentially fruitful; however, their chapter goes no further than this remark, thus leaving the topic of Autocomplete rather unexplored. Finally, two chapters in König and Rasch's edited collection address Autocomplete directly: Martina Mahnke and Emma Uprichard's "Algorithming the Algorithm", and Mary E. Luka and Mél Hogan's "Polluted and Predictive, in 133 Words"; however, both are listed as "creative reflections". This designation is important; all four authors listed are conventional academics, either professors or graduate students, but they chose to frame their research as a series of creative provocations that shy away from any attempt to provide robust or systematic

---

Privacy and the Power of the Algorithm" (Karapapa and Borghi), and "Cache-22: The Fine Line Between Information and Defamation in Google's Autocomplete Function" (Popyer).

conclusions. Hogan and Luka, for example, collected Autocomplete suggestions for a list of 133 queries, chosen at random, over a four-year period; they highlight two, [daughter is] and [boyfriend is], in their discussion:

> I think the words and the searches they suggest point to things people are searching out in private. Real questions. Real worries. It might show how fucked up we all are. What do you think? Look below at "daughter is" – while the suggestions change a little over time, the sentiment remains the same. I find that it's the words that denote a relationship – daughter or boyfriend, for example – are the most deranged. People searching out answers about (and maybe on behalf of) others? Can Google suggestions become a way to suss out what others are experiencing? Is it where we go to feel normal, even if that normal is twisted?

2010	2011	2012	2013
daughter is: overweight depressed moving away pregnant a bully mean tired all the time calling ringtones losing hair a prostitute	**daughter is:** mean pregnant always cold constipated a bully a loner cutting herself tired all the time overweight calling ringtone	**daughter is:** pregnant depressed cutting herself spanish pregnant with dads child constipated mean out of control disrespectful to mother a brat	**daughter is:** pregnant calling ringtone spanish an atheist sexting getting married depressed a brat taller than me a tomboy
boyfriend is: depressed distant ignoring me selfish insecure a virgin moving away a jerk too big	**boyfriend is:** distant depressed selfish sick ignoring me cheap too big gay boring mad at me	**boyfriend is:** distant selfish a virgin boring depressed a jerk sick controlling too clingy insecure	**boyfriend is:** depressed immature a virgin a douchebag boring ignoring me moving away selfish distant gay

(242-243 including table)

Hogan and Luka's discussion touches on a wide range of issues concerning the relationship of individuals to technology. The Autocomplete suggestions that they collect certainly contain a variety of affects that might prove interesting reading, similar to the kinds of decontextualised snippets of conversations one might overhear in public. However, even though Hogan and Luka's lists of query suggestions represent, to my knowledge, the only longitudinal survey of Autocomplete, their chapter provides no deeper an insight into Autocomplete than the YouTube videos of celebrities and teenage vloggers discussed earlier. Their chapter is littered with rhetorical questions regarding why people might search such queries; however, the way in which Autocomplete might operate is untouched. We do not know if these suggestions even relate to queries that are searched; for example, why did the suggestion "[daughter is pregnant with dads child]" (242) only appear in 2012? Do these results suggest a pattern or relationship to wider events? Do these suggestions relate to search volume, or is Autocomplete simply drawing such predictions from a different source? Even if the metrics were clearly outlined, what kind of an analysis could be made of such results? Without an understanding of *why* or *how* Autocomplete suggests the queries it does, no amount of collected Autocomplete suggestions can tell us anything much of value. This difficulty is one of the most pressing aporias facing the study of digital culture; the *how* and *why* of algorithmically controlled interfaces and environments is, potentially, not even something that the software developers necessarily know. For example, the shift towards machine learning, which will be discussed shortly, provides engineers even less control and insight into the outputs of their systems. This chapter represents a drawing together of the available evidence, to present the most probable description of how Autocomplete operates; however, the wider question of access remains

central and the opportunity for critique within digital culture is shrinking, due to the current trend towards blackbox processes and proprietary data, and should not be dismissed as something scholars might overcome. That these processes, central to our global digital landscape, are based upon concealment and require the prerequisites of user ignorance and blind trust, in order to function, should not be ignored. These issues are a central part of Google and the wider digital culture that this thesis addresses in its critique.

There are a large number of reasons why Autocomplete suggestions are important; a number of particular examples will be drawn out throughout this chapter. Can Autocomplete suggestions normalise or encourage certain behaviours? During the 2016 American Presidential election, *Fox News*' chief judicial analyst Andrew Napolitano accused Google of manipulating the Autocomplete suggestions to favour Hilary Clinton over Donald Trump.[4] Some of the biggest lawsuits brought against Google relate to Autocomplete suggestions that build associations between individuals and activities, such as criminal activities or other negative accusations. As will be discussed later in this chapter, such court cases have gone to the highest international courts and encouraged discussion regarding whether individuals and events in their past have a *right to be forgotten*. A large industry of reputation management exists with attempts to game the system of Autocomplete and force out negative suggestions in favour of illegitimate positive ones. These examples will be discussed throughout this chapter and will serve to evidence why Google's tool has a central role in contemporary debates around digital identity, legal freedoms of individuals and the press, normalisation of behaviours, democratic elections, and the praxis of wider social attitudes.

[4] These claims were later deemed inaccurate by fact checking website PolitiFact, see Graves.

2.0 The Origins of Autocomplete

Google Search's Autocomplete tool was invented in 2004 by Google engineer Kevin Gibbs originally as an URL predictor to save time when typing a website address. This built on a much older convention of command-line completion that dates back to the 1960s and which is still common today, whereby programmers use the tab key to complete coding strings used previously in the program. Rather than drawing from previous fixed inputs of a single programmer, however, Gibbs' tool aimed to respond to the open-ended Google queries of all users. Gibbs' release statement introduces the new feature, which at that time was an optional tool, by focusing on two ideas, speed and discovery: "We've found that Google Suggest not only makes it easier to type in your favorite searches (let's face it -- we're all a little lazy), but also gives you a playground to explore what others are searching about, and learn about things you haven't dreamt of" (Gibbs). Autocomplete stayed as an optional "labs" tool until 2008 when it became a compulsory part of Google Search. In one of Google's blog posts, Autocomplete is introduced using an anecdote in which one of the engineers working on the project has forgotten the artist of a particular song, which happens to be a song he wants to be played at his wedding. By typing in some of the lyrics, [from this], Google's Autocomplete tool suggested [from this moment by shania twain] (see fig. 10).

Figure 10: Example of Autocomplete suggestions used by Kevin Gibbs in his 2004 blog post “I’ve got a suggestion”, accessed 05/04/2014.

Without even making a query his question had been answered, Google had helped him remember or perhaps instead the large number of people who had made the same search of [from this moment on by shania twain] had collectively reminded him. The chosen anecdote is deployed to minimise fears: there is only one correct answer to the query and the answer required is unambiguous, there are no risks involved if the answer is wrong and the emotional example of wedding music allows the article to present the change as an unquestionably positive change.

When Autocomplete was made a compulsory feature in 2008, similar appeals to speed were made by Google: “Who wants to spend their time typing [san francisco chronicle] when you can just type in "san f..." and choose the suggestion right away?” (Liu). Also introduced was a sense in which Autocomplete could guide users to formulate different, or at least more specific queries. “Instead of just typing [hotels in washington] - did you want [hotels in washington dc] or [hotels in washington state]?” (Liu). Making users formulate more specific queries simplifies the search process for Google rather than users. Persuading users to search for

[San Francisco Chronicle] articulates their search in terms which can be more easily fulfilled and evaluated from an information retrieval perspective: for example, did this search bring up the newspaper's own website or the website of a competitor, a website describing its history, a Wikipedia page? Users expectations for specific results are statistically easier to predict and navigational queries are easier to satisfy than more open-ended informational ones. The editorial control, via algorithm maintenance, required for a search such as [San Francisco Chronicle] is easier than a query such as [San Francisco], which has a greater number of different contexts or expectations; a short unrefined query of a place might indicate that the user wishes to see an overview of websites relating to San Francisco, or that they have something else in mind but have not included enough information in their query, or that they have no specific intention and are simply curious about the kind of websites that would be returned. A more generalised search for San Francisco places more onus onto Google to make decisions about importance and relevance. What elements should be prioritised, its politics, history, geographical location, culture or recent news? The list of results for such a wide search will never hope to satisfy every searcher; making requests specific increases the chance of success and promotes an attitude that Google's results are objective, simply by virtue of reducing ambiguity. A key consequence, therefore, of Autocomplete is that in addition to changing users' queries at the time of a search, by offering a selection in the drop-down list, it also moulds the expectations used to formulate future queries. Google's other suggested benefit, speed, may well work in combination with such an effect, consequently reducing the time users have to contemplate their queries.

3.0 So How *Does* Autocomplete Operate?

Google have always been very vague about how Autocomplete operates; even though the details of algorithms are not revealed to the public, the general principles behind most aspects of Google's search engine are disclosed by the company. An example of this is Google's openness about how they rank results using the PageRank algorithm; even though the public does not know the details, it is public knowledge that highly ranked pages are those with high numbers of incoming links and/or links from highly ranked pages. An individual without a technical background, or specific knowledge of the PageRank algorithm, can understand that the basic philosophy behind Google's ranking is that it prioritises widely known sites (high hyperlink quantity) and highly trusted sites (sites with hyperlinks from quality sites, such as whitehouse.gov). Understanding this general principle explains why previously established institutions often rank highly in Google's results, or why the public interest in a Wikipedia page correlates with its ranking. Autocomplete has no such general principles and this is why the misapprehension that search volume is the main, or only, criterion has dominated in both public and academic discourse. To understand the extent of Google's vagueness, I have reproduced three versions of their official statement in full, the first is the earliest version of their statement that could be found (18/4/2013, see fig. 11), the second is from 9/4/2015 (see fig. 12) and the third is their current statement, as of 12/8/2017 (see fig. 13). The key recurring phrase that is important for this current analysis is the following: in 2013 "search queries that you see as part of Autocomplete are a reflection of the search activity of all web users and the content of web pages indexed by Google" (Google, "Google Inside Search: Autocomplete"); in 2015 "search queries that you see as part of

Autocomplete are a reflection of the search activity of users and the content of web pages" (Google, "Google Search Help: Autocomplete"); and in 2017 "The algorithm is: Based on several factors, like how often others have searched for a term. Designed to show the range of information on the web" (Google, "Google Search Help: Search Using Autocomplete").

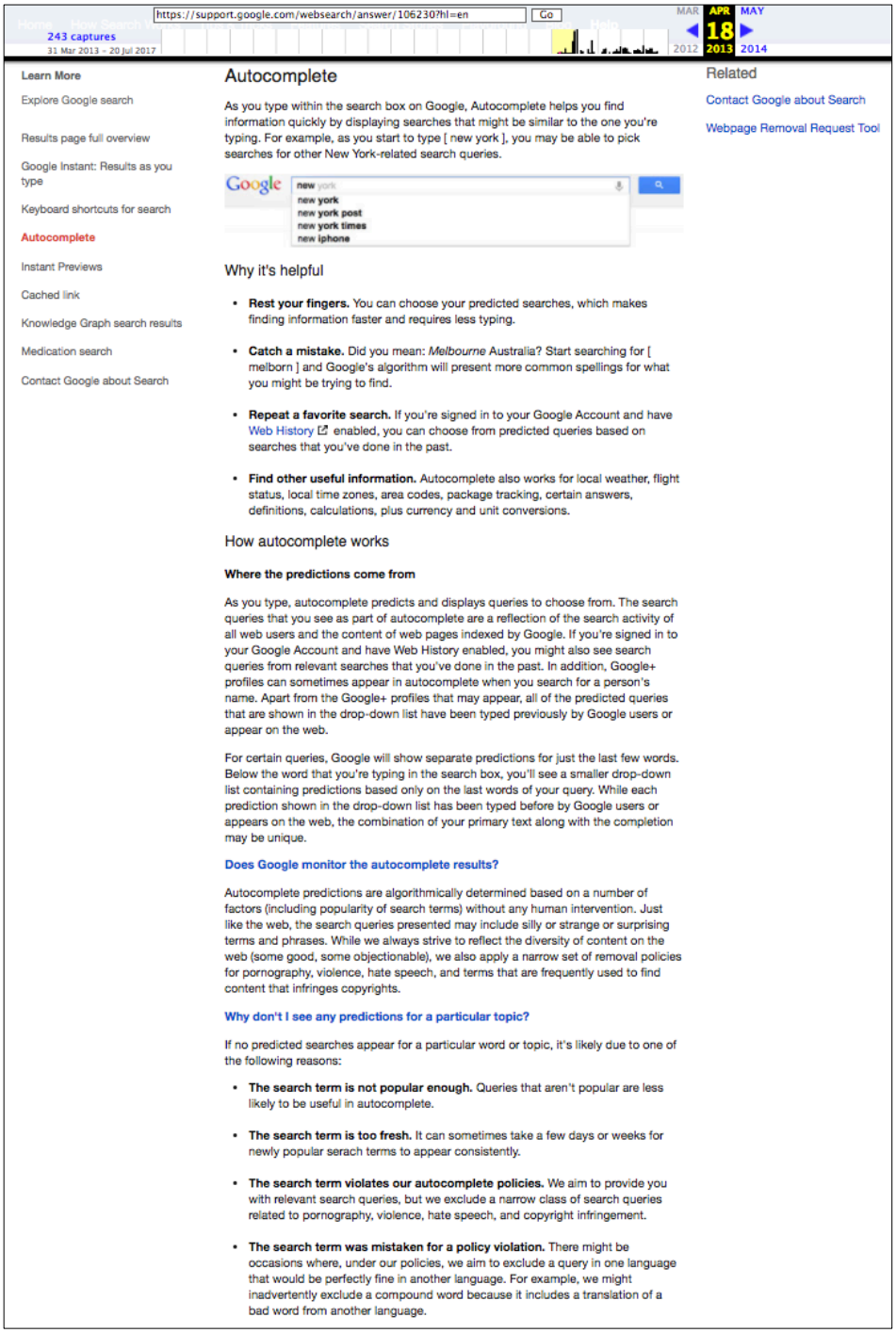


Figure 11. Google's Official Statement regarding Autocomplete "Google Inside Search: Autocomplete" dated 8/4/2013. Accessed 12/8/2017 via the Way Back Machine.

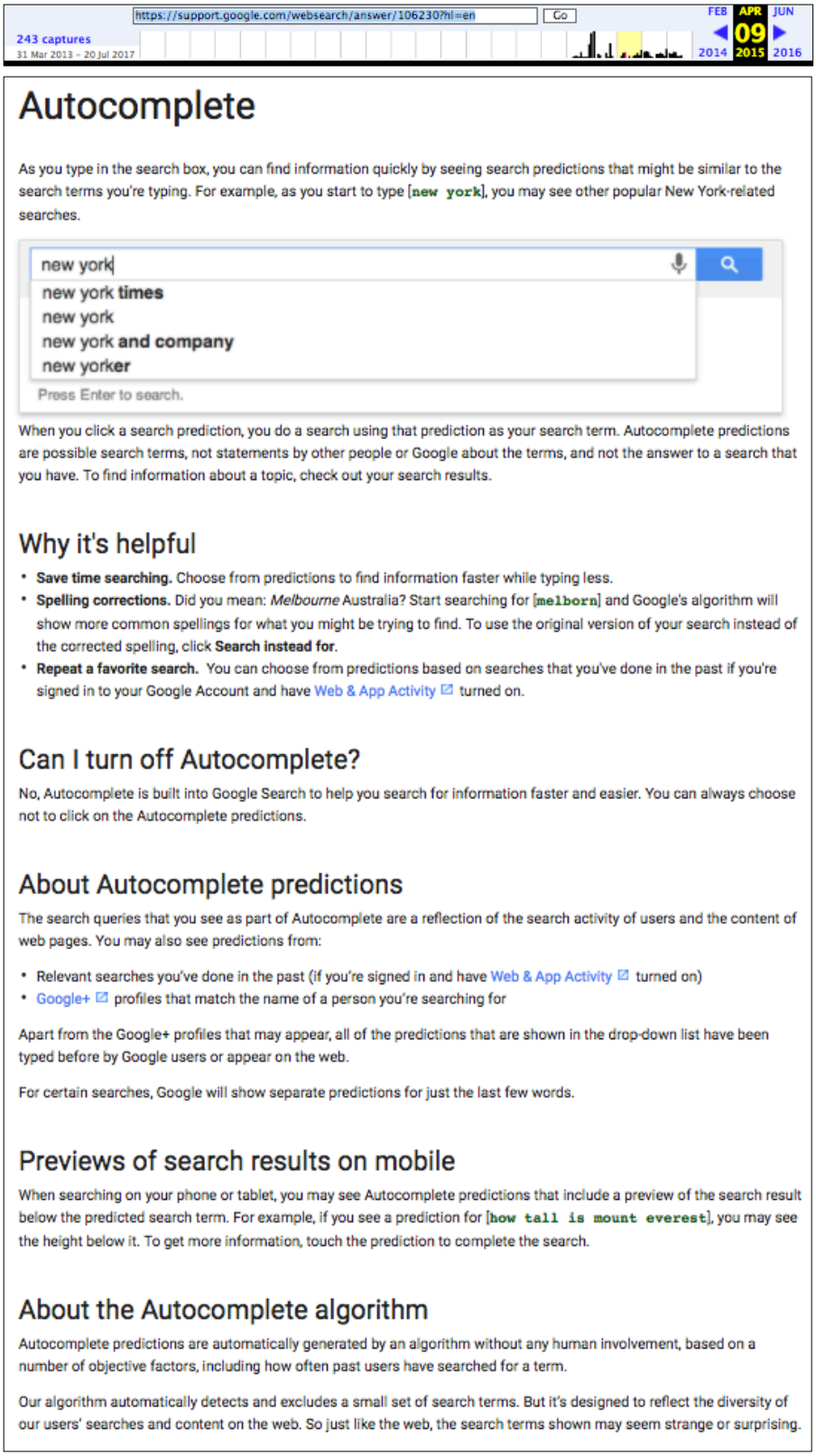
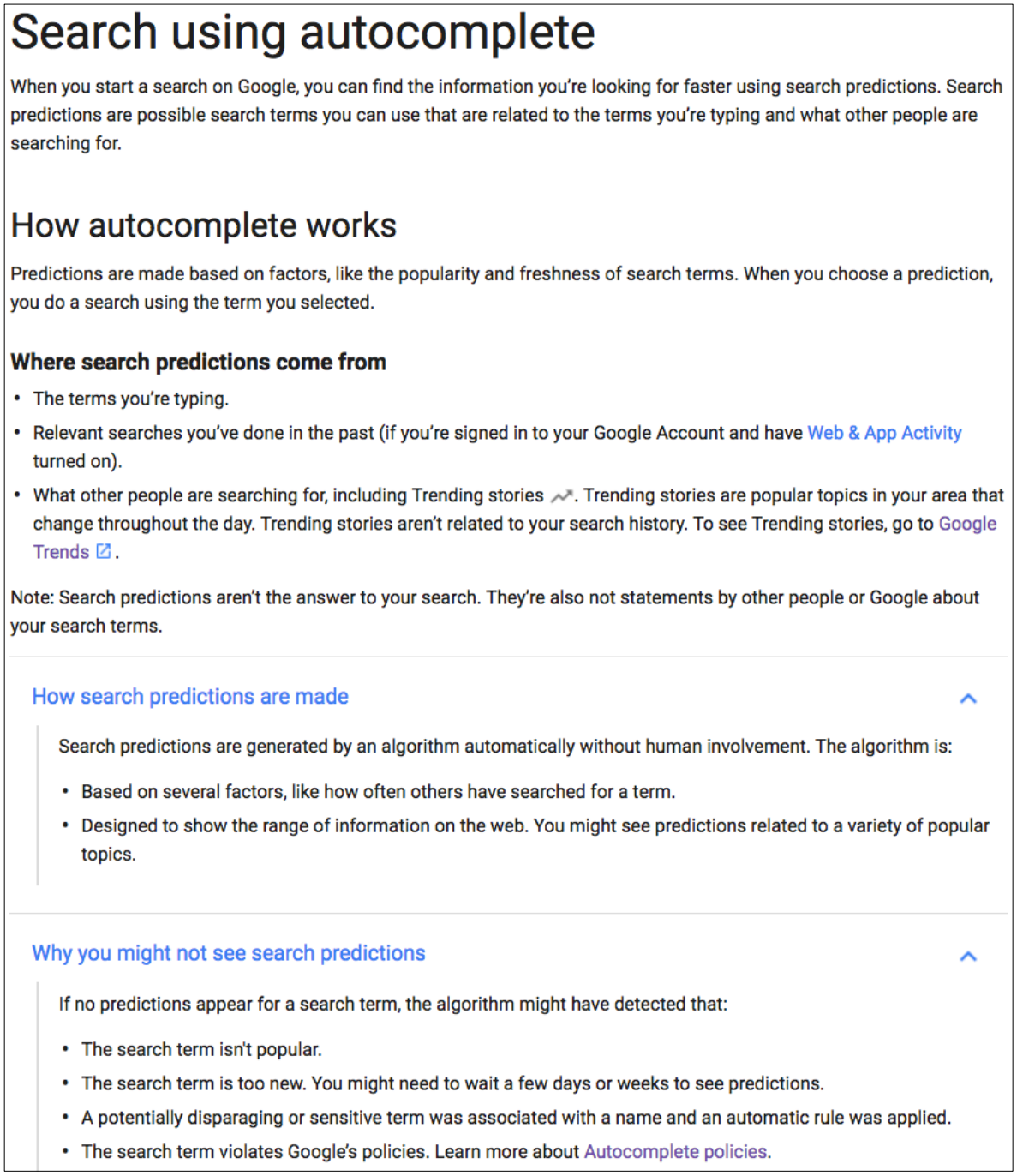


Figure 12. Google's Official Statement regarding Autocomplete "Google Search Help: Autocomplete" captured 9/4/2015. Accessed 12/8/2017 via the Way Back Machine.



Search using autocomplete

When you start a search on Google, you can find the information you're looking for faster using search predictions. Search predictions are possible search terms you can use that are related to the terms you're typing and what other people are searching for.

How autocomplete works

Predictions are made based on factors, like the popularity and freshness of search terms. When you choose a prediction, you do a search using the term you selected.

Where search predictions come from

- The terms you're typing.
- Relevant searches you've done in the past (if you're signed in to your Google Account and have Web & App Activity turned on).
- What other people are searching for, including Trending stories. Trending stories are popular topics in your area that change throughout the day. Trending stories aren't related to your search history. To see Trending stories, go to Google Trends.

Note: Search predictions aren't the answer to your search. They're also not statements by other people or Google about your search terms.

How search predictions are made

Search predictions are generated by an algorithm automatically without human involvement. The algorithm is:

- Based on several factors, like how often others have searched for a term.
- Designed to show the range of information on the web. You might see predictions related to a variety of popular topics.

Why you might not see search predictions

If no predictions appear for a search term, the algorithm might have detected that:

- The search term isn't popular.
- The search term is too new. You might need to wait a few days or weeks to see predictions.
- A potentially disparaging or sensitive term was associated with a name and an automatic rule was applied.
- The search term violates Google's policies. Learn more about Autocomplete policies.

Figure 13. Google's Official Statement regarding Autocomplete "Google Search Help: Search Using Autocomplete" Accessed 12/8/2017.

Comparing figs. 11-13 shows that Google have become increasingly evasive; the key aspect, however, is that there is always a reference to using the content of the web on which to model suggestions, not just the previous queries of other users.

Due to Google's various digitisation projects, in particular Google Books which, as of 2015, had "more than 25 million volumes […] scanned, including texts in 400 languages from more than 100 countries" (Heyman), the content of the web consists of much more than webpages.

In addition, Google have stated that around 15% of all queries are unique and have never been searched before, as Sullivan estimated in 2016: "15 percent of three billion is still a huge number of queries never entered by any human searcher — 450 million per day" (Sullivan, "FAQ: All about the Google RankBrain"). Apart from specifically removed content, Autocomplete is always triggered by a query formulation; Google have outlined that it uses machine learning to produce suggestions for unique queries. We know from Sullivan's "FAQ" that RankBrain, Google's machine learning metric, now triggers for all queries and is the third most important signal for providing results. What is slightly less clear is the extent to which RankBrain provides Autocomplete suggestions. The hypothesis of this chapter is that Google's machine learning has a significant influence on Autocomplete suggestions for all queries, not just the 15% of unique queries. The consequence of this, I argue, is that Autocomplete suggestions might reflect broader textual relationships drawn from a much larger corpus of human thought: books scanned by Google Books, the number of which is steadily increasing towards the inclusion of all written texts. In addition, I will use some examples in the next section to argue that when Autocomplete draws from Google's corpus of previously searched queries during their machine learning process, the statistical probability models produced trigger for individual words and small word grouping patterns, rather than Autocomplete suggestions simply reproducing previous search queries verbatim.

3.1 Probability and Machine Learning

Before exploring the potential extent of Google's machine learning activities, I will briefly outline the general principles of probability that underlie software such as fixed-value Autocomplete technologies and spellcheckers. A basic logic that underlies such systems is still followed even by the more complex machine learning systems that Google employ in various products. Essentially, programming software that predicts or suggests text usually consists of conditional probability model. Two main techniques for calculating probabilities are a trie (short for retrieval, also known as prefix trees) and ternary search trees; these models can be visualised through a diagram such as figs. 14 and 15. A dictionary is used to dictate the array of possible strings, in the case of fig. 14, a dictionary of nine terms leads to the following diagram.

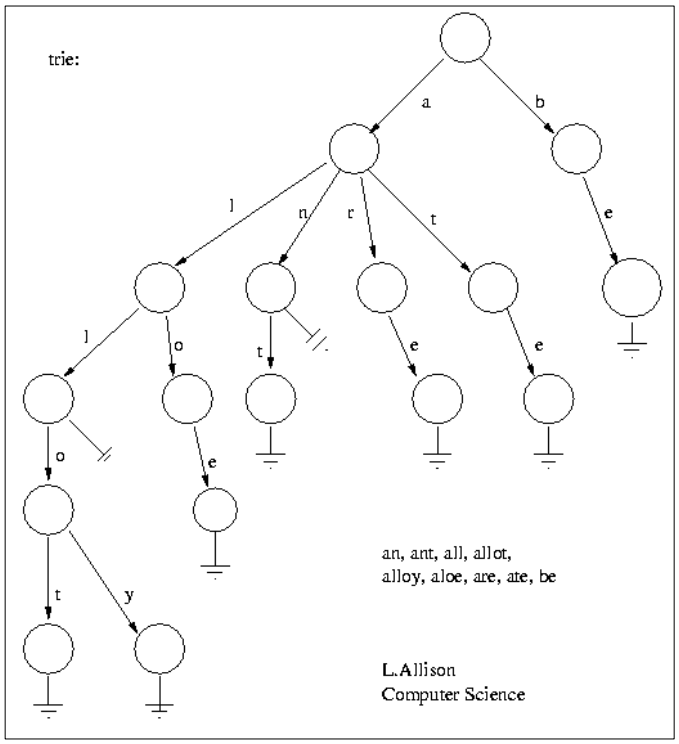


Figure 14. Example of a Trie structure, taken from Lloyd Allison Computing.

A simplistic way of calculating probability is by introducing a corpus and counting the occurrences of strings; fig. 15 demonstrates mapping a count of string occurrences onto a trie diagram.

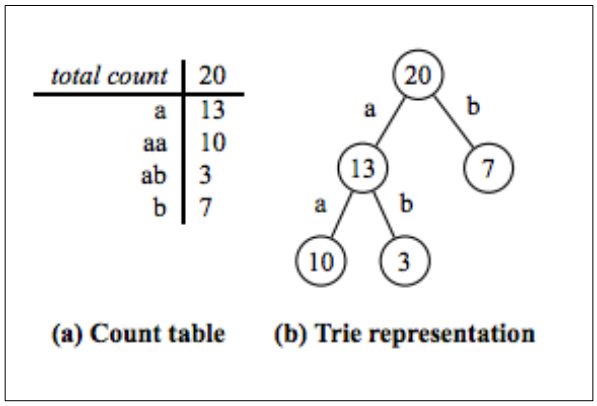


Figure 15. Allocating string counts for an array, represented in a trie diagram, taken from Germann et al. 2009.

This very basic model can be advanced with additional data regarding where the string occurs and, through the use of a Markov Chain (as is employed in PageRank), the most probable strings can be ordered from most likely to least likely, based on the chosen corpus. Writing a program for such an Autocomplete model requires a fairly basic programming acquaintance; for example, Rodrigo Palacios authored a guide on GitHub, titled “Autocomplete - an Adult and Kid Friendly Exercise in Creating a Predictive Program” and explains how to implement such a program in the Python programming language, aimed at the level of a five-year-old. Peter Norvig, previously Google’s director of search quality and currently their director of research, wrote a similar guide, in the Python programming language, for constructing a spelling corrector along the same lines as Google’s spellchecker functions (see Norvig). Due to their simplicity, such simple implementations do not

account for context and grammar; the probability of a word following a particular string needs to take into account factors in addition to word frequency, for example, sentence structure. The more informational tags that can be applied to each word the more accurate the conditional probability measurement will be. For instance, a simple example would be tagging all the words with their grammatical properties by hand, which enables a program to calculate the probability of a verb following a noun as high and the probability of a verb following a verb as low, without giving the program any explicit syntactical rules. Such a system is useful for a small scale Autocomplete program because such additional information is a better predictor than simple letter frequencies. However, due to the scale and variation of the web, Google Search covers a much larger range of language usage and therefore traditional linguistic probability models, outlined above, are not feasible. This has led to the expansion of Google's machine learning neural network program, RankBrain, which has become an increasingly central part of Google's overall company strategy and has significant consequences.

3.2 RankBrain

RankBrain is a machine learning neural network that was originally established to increase the success rates of results provided for unique queries. RankBrain was designed to trigger for all queries that had never been searched before, which represented 15% of all searches submitted to Google, and reword that query into a query that had been searched before. RankBrain would then track whether the user's actions seemed to indicate whether or not this substituted query was successful or not and then record the success or failure as a piece of information

about how similar or different the original query was from its replacement. As Sullivan describes:

> Imagine that RankBrain sees a search for [best flower shop in Los Angeles]. It might understand that this is similar to another search that's perhaps more popular, such as [best LA flower shops]. If so, it might then simply translate the first search behind the scenes into the second one. It would do that because for a more popular search, Google has much more user data that helps it feel more confident about the quality of the results. ("Google Uses RankBrain for Every Search")

This system is much more advanced than simply assigning a range of synonyms to each word, as had previously been initiated. Due success of RankBrain, Google started to employ the machine learning neural network for all queries and as reported in Sullivan's "FAQ: All About the Google RankBrain Algorithm" it is now considered Google's "third most important signal" after "content" and "links": the bedrock of search engine logic. Considering that Google use around 200 signals broken up into "10,000 variations or sub-signals", RankBrain does not simply represent an issue related to Autocomplete, but constitutes a key part of Google's whole system. In addition, Google's competitors also rely on machine learning; Microsoft, for example, has been working on its machine learning system, RankNet, since 2005, employing it as part of Bing's ranking signals.

In 2011, Google started training its machine learning software on their Google Books corpus, which at that time was around "6% of all books ever published" (Lin et al.), focusing on books in eight languages from 1800 to the present. Google

researchers document, in a number of academic papers, their process for using machine learning to syntactically annotate n-grams for those eight languages, i.e. automatically assigning words grammatical values based on their use throughout their Google Books. In addition, RankBrain was also able to map the changing usage from 1800 to the present. The kinds of language use found in the Google Books corpus does not necessarily reflect patterns of online language use. In chapter one, the distinction between questions and queries gave an insight into how search queries follow patterns, but not necessarily grammatically correct ones. For this reason, Google's RankBrain system is also trained on the language of queries submitted to Google, as well as the born-digital content of the web.

The machine learning behind RankBrain functions by building vectors from corpuses of text. Fig. 16, taken from one of Google's academic papers written by Mikolov et al. "Learning the Meaning Behind Words" demonstrates these vectors through a Principal Component Analysis (PCA) projection.

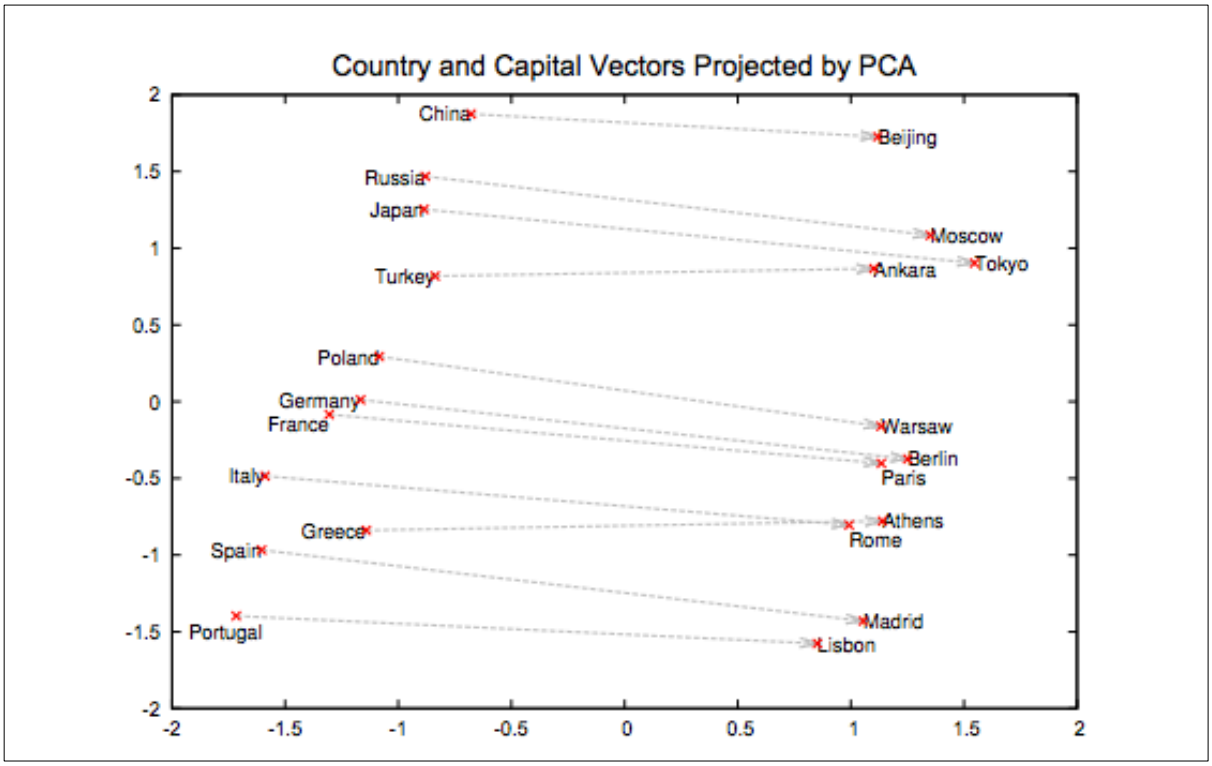


Figure 16. taken from Mikolov et al. “Learning the Meaning Behind Words” 2013 original caption reads: “Two-dimensional PCA projection of the 1000-dimensional Skip-gram vectors of countries and their capital cities. The figure illustrates the ability of the model to automatically organize concepts and learn implicitly the relationships between them, as during the training we did not provide any supervised information about what a capital city means”.

Using a system called Skip-gram, the machine learning system builds a neural network, which “can train on more than 100 billion words in one day” (Mikolov et al., “Distributed Representations” 2) In doing so, the system builds up a series of relationships between words and groups of words (n-grams). Such a system can operate in tandem with Google’s Semantic Search methods discussed earlier to assign values to search terms so that they can be tagged as people or places and be treated differently computationally. Figure 16 shows a simple example of the neural

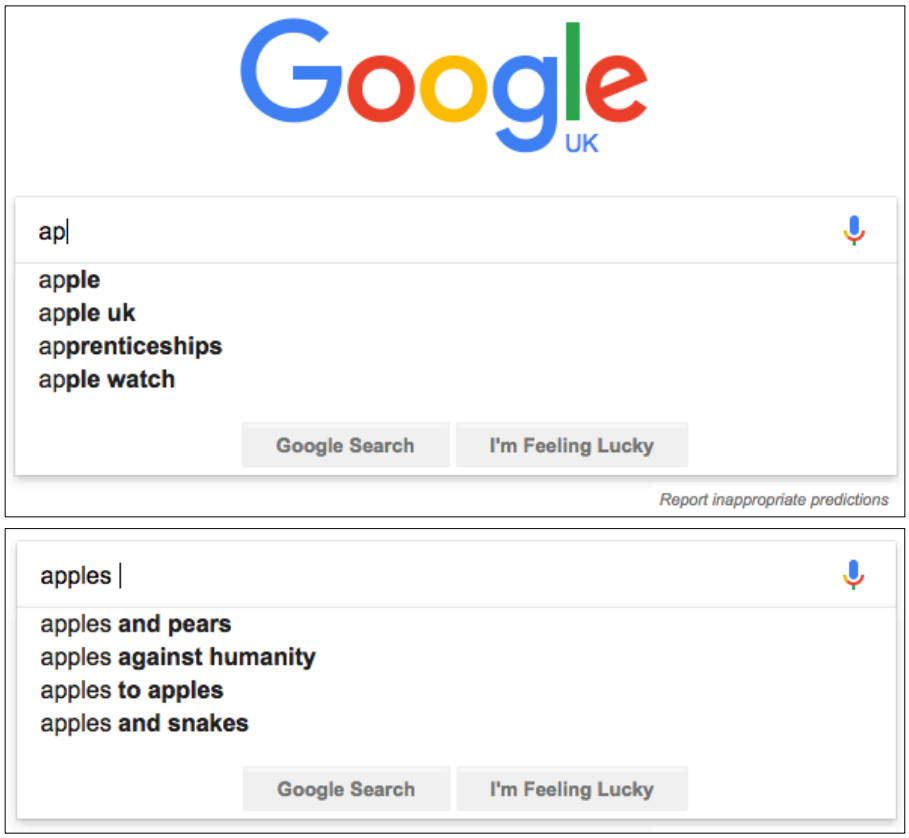
network's organisational capacity to recognise the relationships between capital cities and countries, *and* the different relationship between capital cities and other capital cities, which it plots as vectors. Such a model is useful for idiomatic language or instances where combinations of words are frequently used together to signify a separate concept, Mikolov et al. use the example that "'Boston Globe' is a newspaper, and so it is not a natural combination of the meanings of 'Boston' and 'Globe'. Therefore, using vectors to represent the whole phrases makes the Skip-gram model considerably more expressive" (2). Google's use of machine learning, often described as AI, is another nexus in which the disparate parts of Google's company (and the wider projects of Alphabet) become related. Google have deployed machine learning in a number of different contexts and provides a further insight into how many of their projects are deeply related. Google now use machine learning in all their products, from Google Maps' Street View to YouTube's method of suggesting videos to users. Steven Levy described machine learning as Google's core "corporate mindset" citing CEO Sundar Pichai's statement in late-2015:

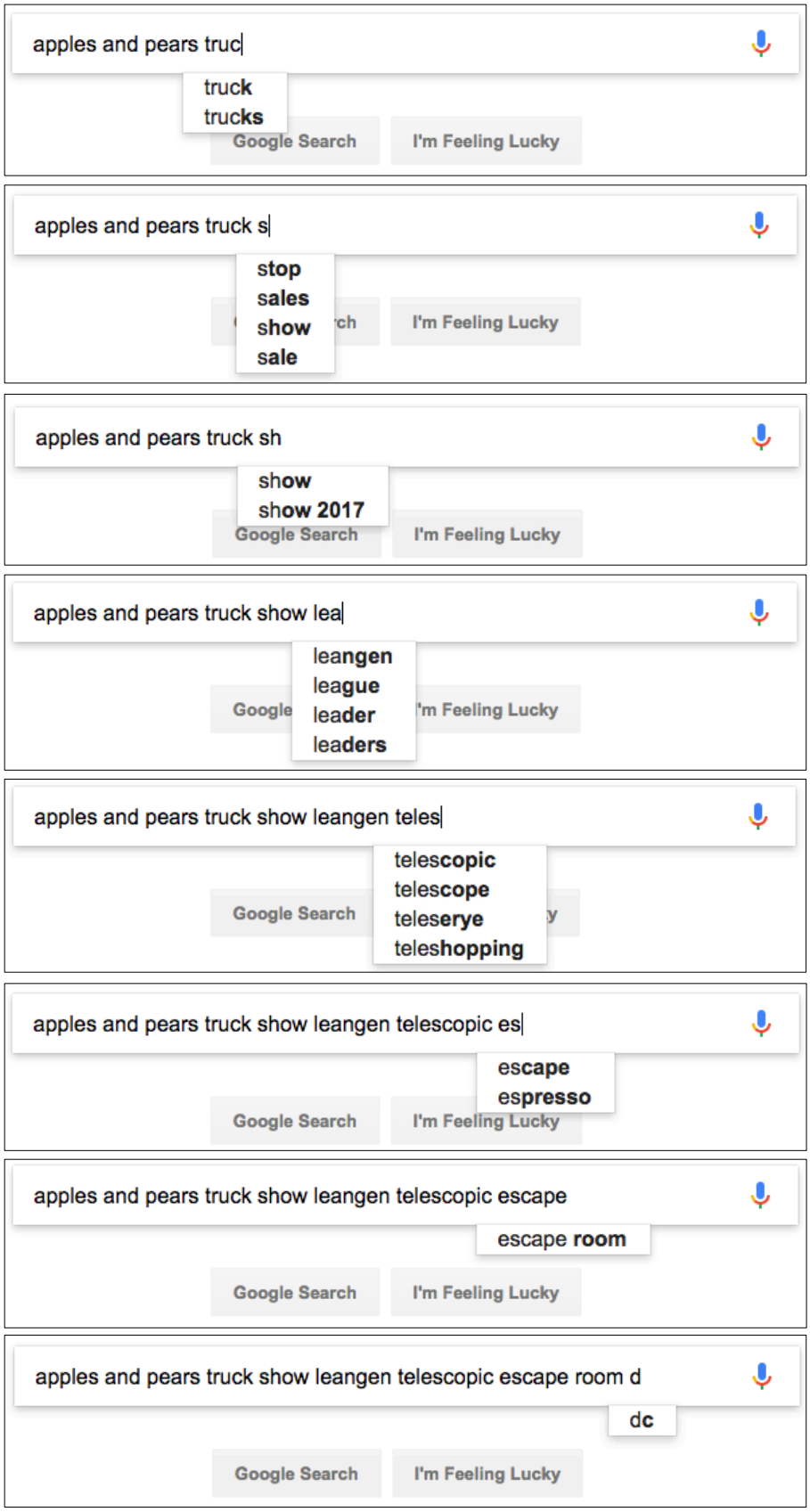
> Machine learning is a core, transformative way by which we're rethinking how we're doing everything. We are thoughtfully applying it across all our products, be it search, ads, YouTube, or Play. And we're in early days, but you will see us — in a systematic way — apply machine learning in all these areas. (qtd. in Levy "How Google is Remaking Itself as a 'Machine Learning First' Company")

Although this thesis is focused on Google's search engine, that technology is part of a much wider narrative: scanning all the books ever published, photographing every

street, collecting data on offline behaviours in people's homes, aggregating data of every dynamic pattern it can from influenza epidemics to traffic flow, funding companies such as 23andme, a service for reading the DNA of customers, all of these and more, represent the logic, in which Google see the world as data to be collected and put to use. In particular, the collection of data in such large quantities enables and reinforces the use of neural network machine learning. At the start of the Google's book scanning project, which aimed to photograph all pages ever printed, an engineer summed up the project in the following way: "we are not scanning all those books to be read by people. We are scanning them to be read by an AI" (Uncredited Google Engineer qtd. in Dyson 313).

The way in which the logic of machine learning leads the Autocomplete process can be seen through the activity of formulating nonsensical queries (see figs. 17-26).





Figures 17- 26: Following Autocomplete suggestions to formulate a nonsensical query. Formulated 12/8/2017.

The above example is a far cry from the Autocomplete questions answered by celebrities and vloggers that opened this chapter. In formulating the query, I did not necessarily follow the first link, instead, I explored the range of options provided by Autocomplete. There is a logic that can be discerned from the pattern, words that correspond to one another: [apples and pears], [truck show], [escape room]. There is also a kind of syntax provided by the types of information suggested: dates [2017], locations [leangen], [dc]. However, the overall arrangement is incomprehensible, [telescopic espresso]? Cade Metz, in a 2016 *Wired* article, discusses the shift in Google's management structure after the retirement of Amit Singhal, then Head of Google Search, and replacement by John Giannandrea. Metz describes the transition as, an "ideal metaphor for a momentous shift in the way things work inside Google – and across the tech world as a whole", that is, from algorithms that followed a strict set or rules coded by engineers to machine learning in the form of neural networks and forms of AI. Metz cites a number of Google's current and ex-employees including software engineer Edmond Lau, who argues that

> Singhal carried a philosophical bias against machine learning. With machine learning, he wrote, the trouble was that "it's hard to explain and ascertain why a particular search result ranks more highly than another result for a given query." And, he added: "It's difficult to directly tweak a machine learning-based system to boost the importance of certain signals over others." Other ex-Googlers agreed with this characterization […] Yes, Google's search engine was always driven by algorithms that automatically generate a response to each query. But these algorithms amounted to a set of definite

> rules. Google engineers could readily change and refine these rules. And unlike neural nets, these algorithms didn't learn on their own. As Lau put it: "Rule-based scoring metrics, while still complex, provide a greater opportunity for engineers to directly tweak weights in specific situations." (qtd. in Metz)

Metz's article was published in February 2016, at this time RankBrain, Google's machine learning neural network was running on the 15% of unique queries, as outlined earlier. However, it only took until June 2016, as documented by Sullivan, for RankBrain to become Google Search's third most important metric, activating for every search. Chapter four includes a comparison between search results before this shift, November 2015, and after the shift, 2017. From the small sample of results collected, it is clear that queries submitted after June 2016 produced results that were highly consistent across different contexts. In particular, queries that used very different phrasing and even different languages resulted in many commonly occurring results. This suggests that Google is not simply treating each word or phrase as an isolated query, instead, it seems that the results fit within a wider structure of presumed function. Such patterns of organisation provide evidence that RankBrain's status as the third most important metric has a concrete impact on the nature of search results.

Google have been using similar AI neural network technology in a number of their ventures, since at least 2011, when Google Brain was founded. Andrew Ng, founder and former head of Google Brain, Google's deep learning neural network programme, gave a sense of its pervasiveness in a 2012 interview with NPR:

> Machine learning and artificial intelligence is a pervasive technology today, and most of us use it dozens of times a day without knowing it. Artificial intelligence technology is responsible for giving us high-quality Web search engines, practical speech recognition, machine translation, even self-driving cars.

In the interview, Ng explains how Google Brain gave a neural network of 16,000 computer processors “three days on 10 million YouTube clips” without explicit instructions or directions. In a conclusion that feeds neatly into the soundbite culture prevalent in technology journalism, after the three days were up the neural network “learned how to recognize a cat”. The nuance of what such a claim really means is often glossed over. A later Google project, DeepDream, developed in 2014, “was invented to help scientists and engineers to see what a deep neural network is seeing when it is looking in a given image” (Google, “DeepDream: About”). In *The Atlantic* article “When Robots Hallucinate” Adrienne Lafrance demonstrates how DeepDream interprets an image (see figs. 27 and 28)

Figure 27. An image given to DeepDream to interpret. Taken from Adrienne Lafrance's "When Robots Hallucinate".

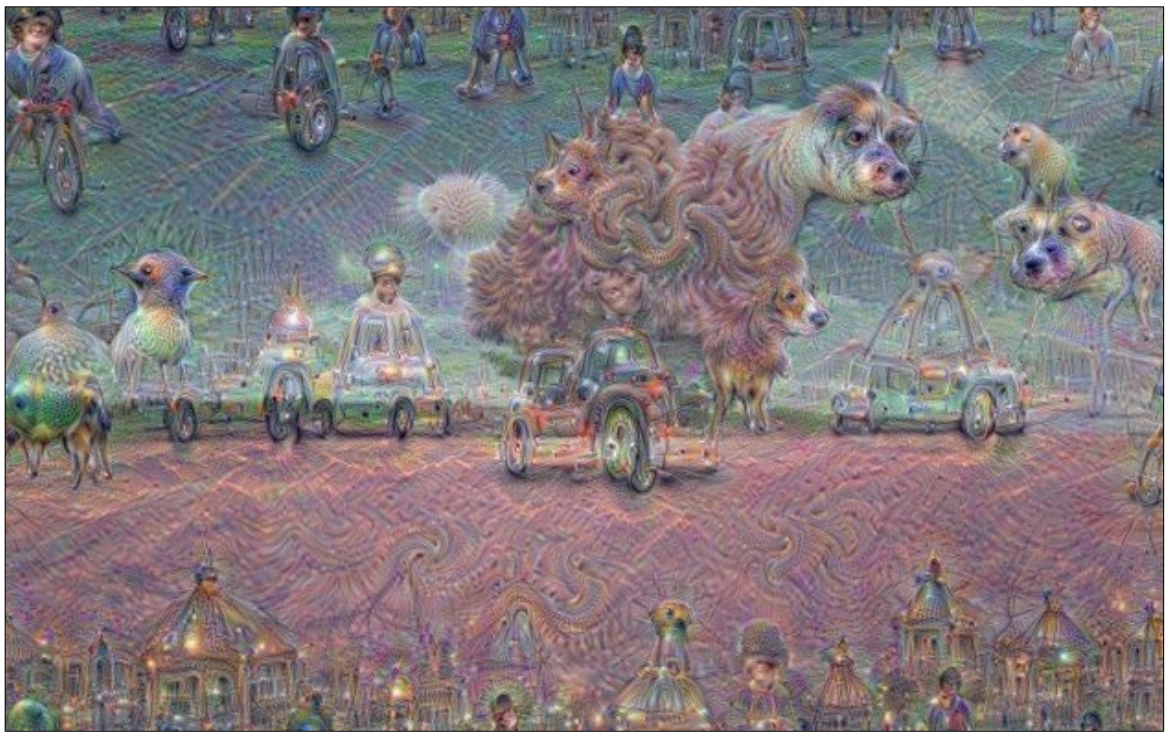


Figure 28. The image of fig. 27 processed by DeepDream. Taken from Adrienne Lafrance's "When Robots Hallucinate.

These images give some insight into the kinds of pattern-matching systems developed in neural networks, but do little more than play into the fantasies inherited from science fiction, literarily complete with androids dreaming of electric sheep. However, to return to Metz's survey of current and ex-Google employees, the difference between what it means to *understand* the processes of algorithms versus neural networks is significant:

> The truth is that even the experts don't completely understand how neural nets work. But they do work. If you feed enough photos of a platypus into a neural net, it can learn to identify a platypus. If you show it enough computer malware code, it can learn to recognize a virus. If you give it enough raw language—words or phrases that people might type into a search engine—it can learn to understand search queries and help respond to them. […] As Google moves search to this AI model, it's unclear how the move will affect its ability to defend its search results against claims of unfairness or change the results in the face of complaints. (Metz)

Metz argues that even experts cannot fully understand how neural networks work, which is troubling in itself. However, more significant is the binary attitude whereby such techniques can work or fail to work, as if search engines can be described in the same terms as a program that can recognise, or fail to recognise, a virus. Relying on the word *work* conflates two aspects of Autocomplete: how Autocomplete *operates*, colloquially what lies under-the-hood, and how Autocomplete *functions*, the social and technological consequences of its implementation. As emphasised throughout this thesis, algorithms are never neutral and always codify a set of

cultural, social, and philosophical attitudes. The implementation of neural networks, even if their operation is more opaque, does the same. The operational differences between algorithms and neural networks are significant, but their functional outcomes are intertwined. However, knowing that they operate in tandem, but with little evidence concerning their level of influence, means that a critique of either logic is difficult. This exacerbates the problems caused by the hidden nature of Google's ranking criteria and has particular ramifications for Google's accountability for its results and suggestions, both in the legal context of potential libel and in a social context regarding the perpetuation of stereotypes. The following section addresses an example that highlights these issues as well as the consequences of the widespread misunderstanding of how the tool operates.

4.0 The Case of Tuuli Lappalainen

On the 15th of July 2017, the genomics scientist Tuuli Lappalainen tweeted the following screenshot (fig. 29) showing the Google Autocomplete suggestions for her name:

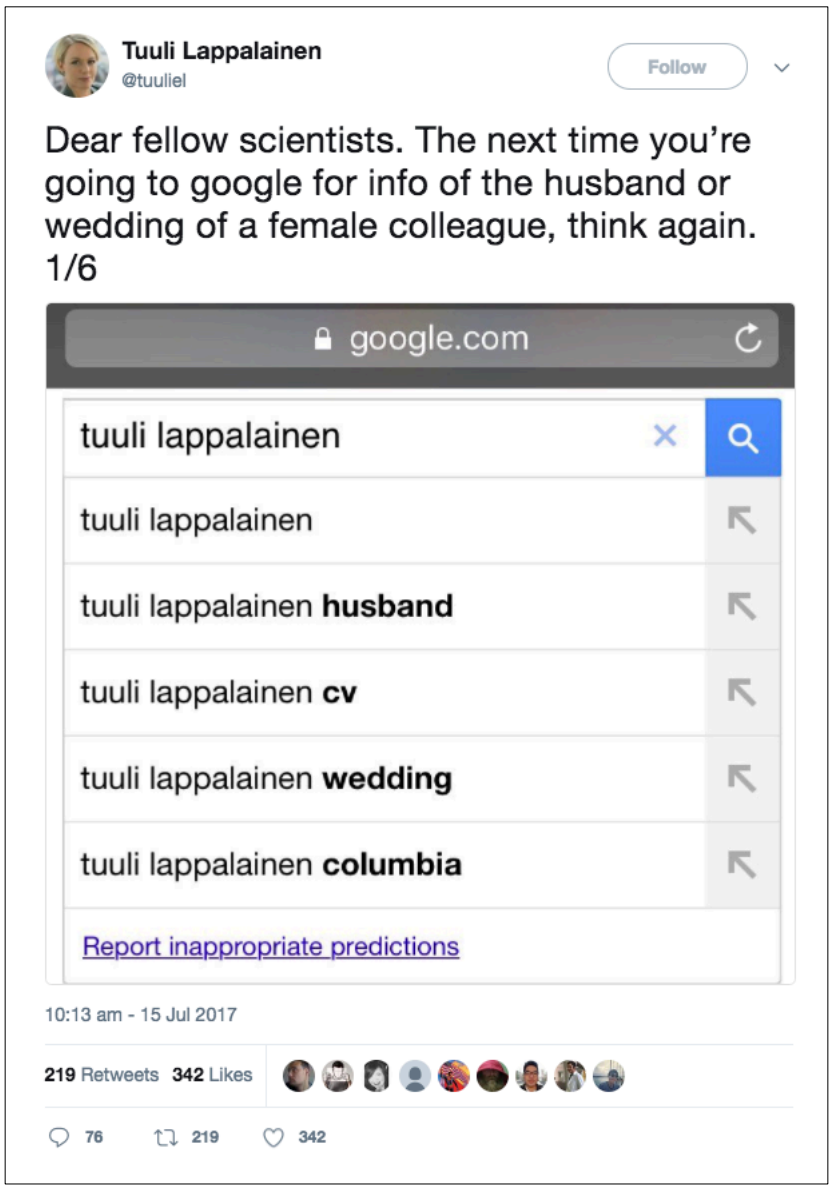


Figure 29. Tweet from Tuuli Lappalainen showing a screenshot of the Google Autocomplete suggestions for her name. Posted to Twitter 15/07/2017.

She continued her statement over five following tweets (fig. 30) which were as follows:

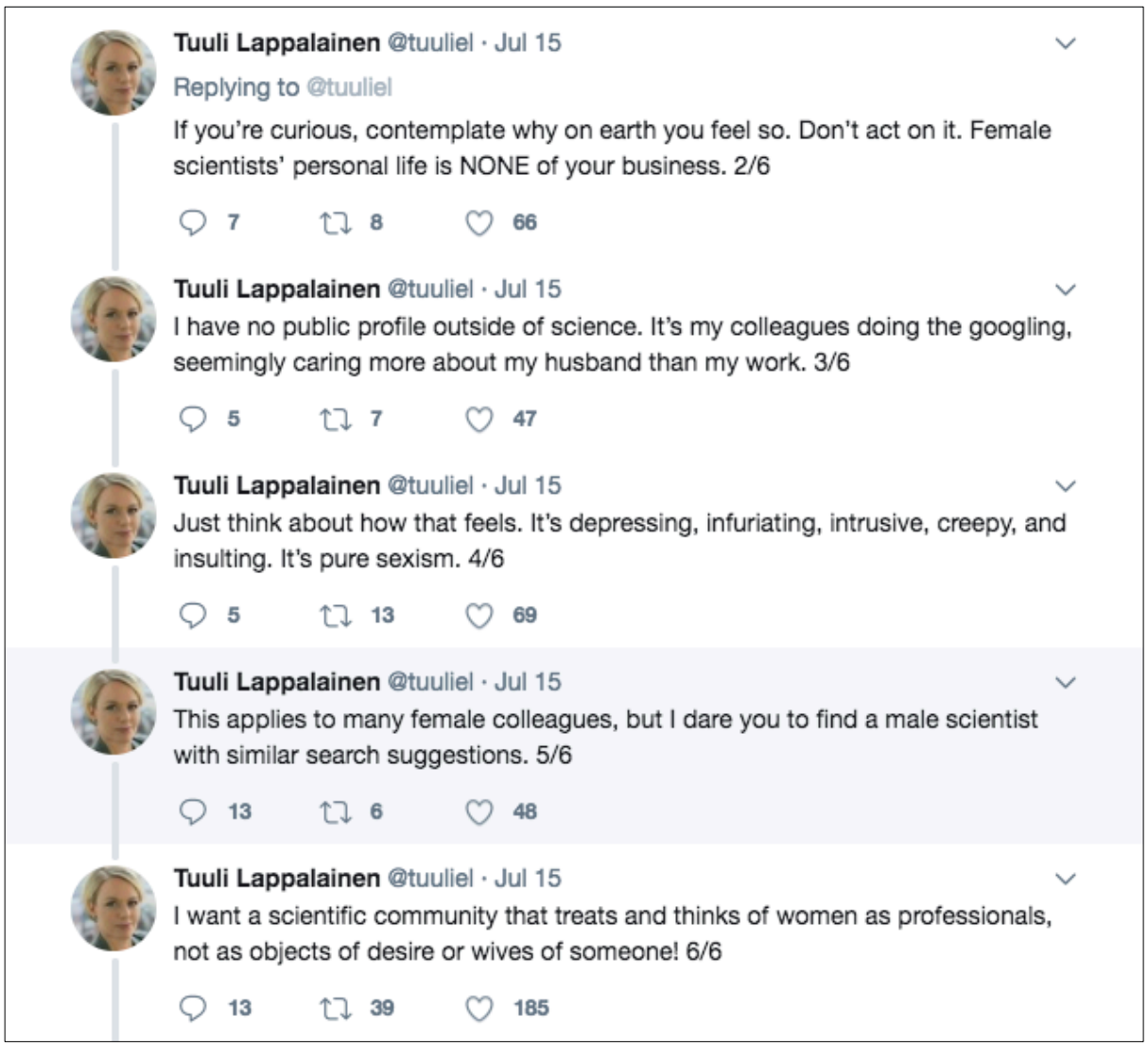


Figure 30. Five further tweets from Tuuli Lappalainen in "reply" to her original screenshot (fig. 29). Posted to Twitter 15/07/2017.

First of all, Lappalainen's comments reflect the erroneous, but widespread, attitude that Autocomplete suggestions are an exclusive reflection of search volume, here the mischaracterisation has led to an outrage focused on the presumption of search patterns. Consequently, Lappalainen's tweets became a site of discussion in which various Twitter users weighed in with a range of comments and examples. Users

retweeted previous examples found by neurobiologist Leslie Vosshall in 2016 (see fig. 31).

Figure 31. Tweet from @pollyp1 (Leslie Vosshall), originally from 20 May 2016. Retweeted by @Scitabanis 15/07/2017.

The anecdotal findings of Vosshall in 2016 led evolutionary ecologist Florence Débarre to collect and analyse the Autocomplete suggestions for a range of professions: scientists, tennis players and Hollywood actors and “considered the proportions of people for whom ‘husband’ and/or ‘wife’ are suggested; [and] also

looked for the word ‘married’”. The results were published in a blog post in which Débarre acknowledges the difficulties of reproducibility, in particular, the impact of location on results. Débarre’s scientists were taken from those listed in the American Academy of Arts & Sciences, Howard Hughes Medical Institute, and TED talk speakers listed under science. The TED talk aspect was stressed in an attempt to evaluate the impact that relative fame might have. Débarre’s findings were that “the proportions of ‘wife’ and/or ‘husband’ [suggested by Autocomplete] are higher among Google suggestions for female scientists” (see fig. 32). Débarre also ruled out the hypothesis that relative fame was associated with suggestions of a personal nature, such as ‘husband’ or ‘married’. The data showed no examples of male scientist TED speakers who had ‘wife’ or ‘husband’ listed in their Autocomplete suggestions, whereas there were a number for female scientists, although this number was significantly lower than for female scientists who had not given a TED talk. Therefore, such personal suggestions could not be attributed to the consequences of being in the public eye and being searched for in a non-academic context, as was regularly suggested in the wider Twitter debate.

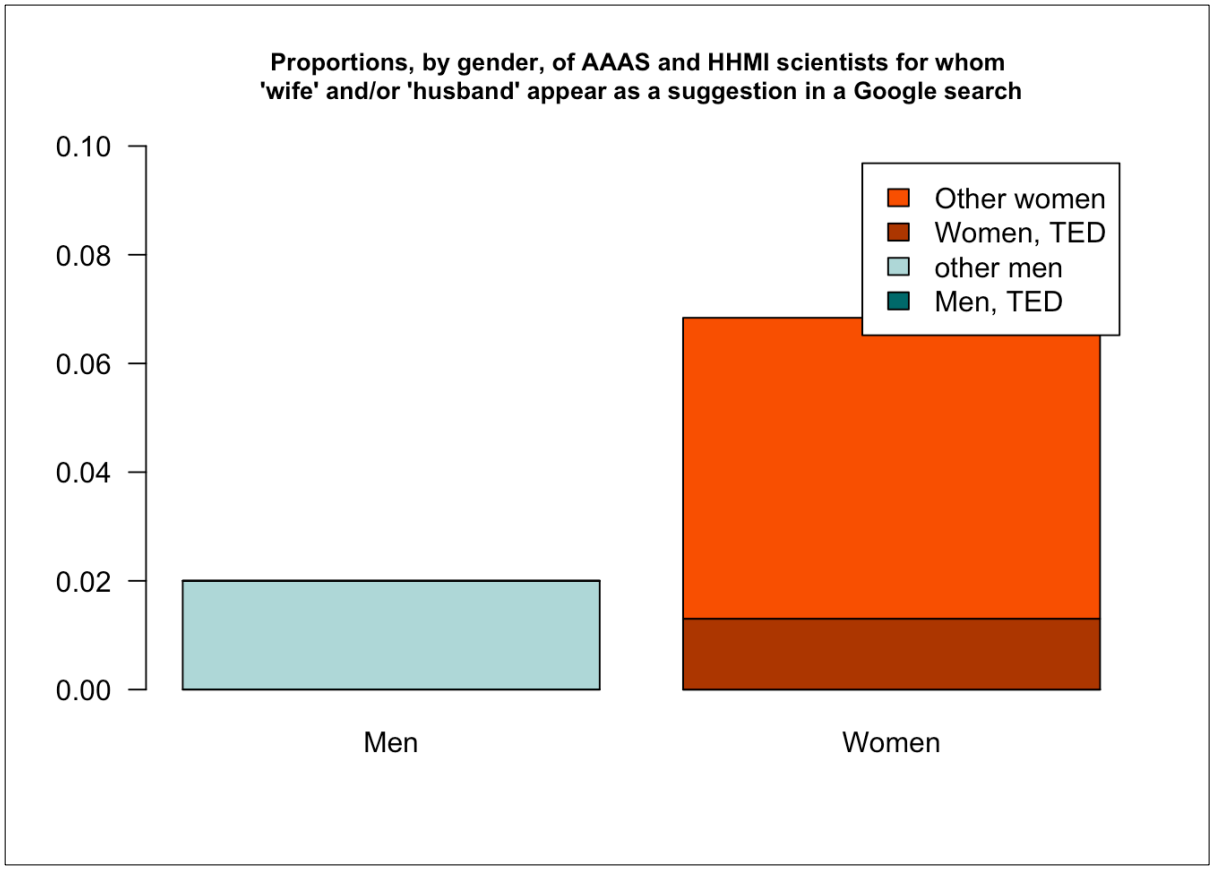


Figure 32. Proportions, by gender, of AAAS and HHMI scientists for whom ‘wife’ and/or ‘husband’ appear as a suggestion in a Google Search. Taken from Florence Débarre’s “Are Google Suggestions Sexist?” This graph shows that relative fame, as represented by an individual having given a TED talk, does not have a significant impact on the gender distribution of the results.

These results confirm the anecdotal reports, presented on Twitter, that keywords such as ‘wife’ or ‘husband’ occur more often in searches for female scientists than for male scientists. The level of robustness and reproducibility of such findings is, however, complex. As outlined throughout this thesis, studying Google’s search engine provides a number of challenges: we do not know the extent to which the results are personalised, each set of searches provides further data that will alter the

algorithmic analysis, and Google as a company constantly change the algorithms involved in response to various factors. With those issues in mind, a Google search for Tuuli Lappalainen on 14/8/17, using a virtual private network (VPN) in order to depersonalise suggestions as much as possible, resulted in the following Autocomplete suggestions (see fig. 33):

Figure 33. Google Autocomplete suggestions for Tuuli Lappalainen, accessed 14/8/2017.

It is impossible to know whether the highlighted suggestions are being caused by Autocomplete algorithms or by Google's RankBrain machine learning neural network. The examples provided by Vosshall and Débarre occurred in the summer of 2016; it was around this time that Google announced that RankBrain was no longer exclusively focused on unique queries and instead had become the third most important ranking signal for all queries. Either way, there is some evidence that the Autocomplete suggestions for specific people is gendered en masse. Research scientists may well represent a case study that reflects the original requirements for

RankBrain, a large selection of queries (in this case individuals) that individually have small search volumes (the consequence of which is a small amount of data for each query), but that share commonalities. How such a commonality is established is an important issue and its consequences bring a number of difficult debates regarding representation. Are such suggestions based on the aggregated data of previous searches, or do they represent the more historical gendered relationships that would be evident in the Google Books corpus? In addition, is the fact that these Autocomplete suggestions continue to occur, even after direct feedback[5] a sign that such suggestions are deemed useful to a majority of users?

What makes the examples above particularly interesting is that suggestions are directed towards specific individuals, rather than widespread groups. The issue of Autocomplete's proclivity towards stereotypes has been highlighted in the popular press in a number of contexts, but in these examples, the stereotypes are attached to common nouns of gender, nationality, ethnicity or religion rather than the proper nouns of individuals. An example is highlighted by a 2013 campaign by UN Women, which concerns the Autocomplete suggestions associated with the word women. The UN Women's advertising campaign used "genuine Google searches to reveal the widespread prevalence of sexism and discrimination against women. Based on searches dated 9 March 2013, the ads expose negative sentiments ranging from stereotyping as well as outright denial of women's rights" (UN Women). The advertisements (see figs. 34-37) depicted the Autocomplete suggestions for the particular phrases entered: [women cannot], [women shouldn't], [women should] and [women need to] all of which portray highly sexist attitudes.

[5] The Twitter threads, previously highlighted, contain messages from a number of users stating that they had reported the suggestions using the associated "report inappropriate predictions" link.

Figures 34-37. UN Women Campaign highlighting the Autocomplete suggestions for [Women shouldn't], [women should], [Women cannot], [Women need to] See "UN Women ad series reveals widespread sexism" 21/10/2013. Accessed 18/05/2014.

The UN Women campaign gained a wide media attention and spawned a number of copycat campaigns for other issues, including examples relating to race, religion, and nationality that showed negative stereotyping (see figs. 38 and 39).

Figure 38. "Racism. It Stops With Me" 2013 Campaign. Featured in Marc van Gurp's "Google's Autocomplete Feature Is Shocking Too About Black Men".

Figure 39. Original caption: “Also try” taken from Marc van Gurp’s “The Shocking Answers from Google’s Autocomplete Feature on Sexism”.

These results are morally outrageous, but where are we to place the blame? As outlined above, Autocomplete suggestions are drawn from the linguistic arrangements of search patterns and the content of the web, which includes the (as of 2015) 25 million volumes of printed text scanned by Google Books. The lack of detail provided by Google regarding how these different sources are weighted makes it impossible to ascertain where such attitudes, misogynistic or otherwise, have originated. Autocomplete suggestions are the consequence of a wider linguistic network that, for example, has strong associative links between female names and words related to weddings and marriages. In these examples, Autocomplete becomes a tool: not one that helps users to search faster but a tool that helps highlight a set of attitudes. However, Google’s response to the various campaigns that concerned general nouns was to disable Autocomplete suggestions for those

and related searches. Such a decision may reduce the extent to which certain attitudes are perpetuated, but it also hides the underlying issues that, given their priority in previous suggestions, still need to be challenged. Also, given that such attitudes are still perpetuated in the less explicit examples, whereby attitudes attached to personal names or other single instances only become clear in aggregate, the action of removing specific highlighted examples does not address the deeper issue that Autocomplete suggestions will always represent and reiterate a set of values. These examples show that the neural networks have learned about gendered discourse and that, in the main, these suggestions have been deemed *useful*: users have clicked on them, which perpetuates the problem. At the beginning of chapter one of this thesis, a quotation from Bruno Latour's essay "Technology Is Society Made Durable" was used to highlight how, even at the lowest and most basic level, the programming that underlies Google's search engine reproduces particular cultural values. The same statement is just as applicable to the most complex kind of computing undertaken by Google: neural network machine learning. Latour writes that: "we might call technology the moment when social assemblages gain stability by aligning actors and observers. Society and technology are not two ontologically distinct entities but more like phases of the same essential action" (129). In chapter one, this quotation was used to show that the underlying code of the web, which dictates how sites can be crawled by spiders, represented the social attitudes towards Google as a company and an alignment of "actors and observers". Here, the actors and observers represent a much broader group of individuals than webmasters and programmers, including all online users and the words, ideas and sentiments contained within books scanned through the Google Books project. The discourse around machine learning relies on a lack of human guidance to present

the process as neutral or objective. However, the above examples show that particular values and attitudes are perpetuated and encouraged, or to use Latour's terminology, stabilised and made durable. Google represents a monopoly in many countries globally; consequently, its widespread usage provides its neural network with more data to interpret, reframe and fix certain attitudes.

A related example occurred two years before the UN Women advertisement campaign in a campaign to "sweeten up the Romanian image" (see van Gurp "Romanians are Smart"). The campaign was a reaction against the negative Autocomplete results for [Romanians are] (see fig. 40) and therefore in order to raise the profile of their country Romanians were urged to repeatedly search [Romanians are smart], which the website of the organisation would semi-automate (see fig. 41).

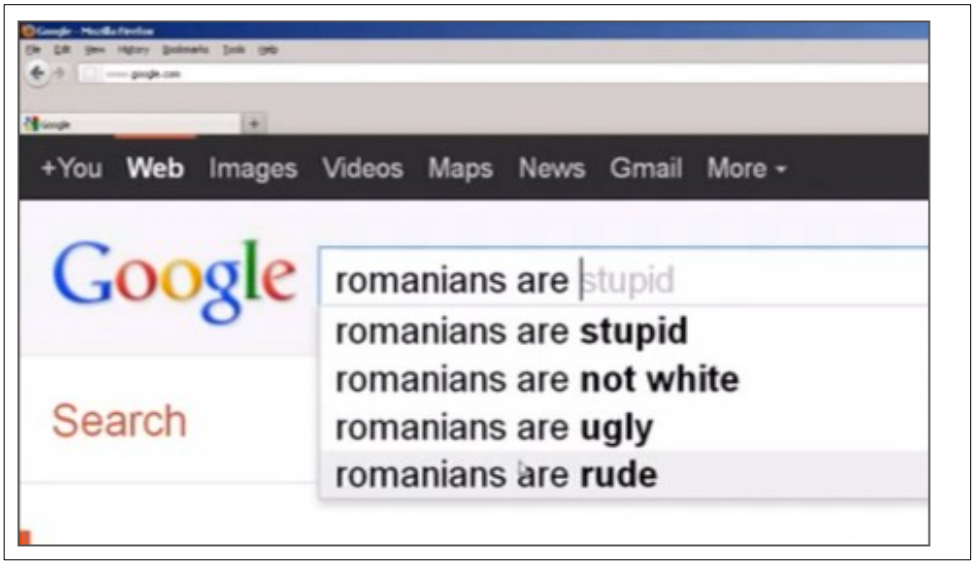


Figure 40. Depiction of the Autocomplete suggestions for [Romanians are], taken from Marc van Gurp "Romanians are Smart".

Figure 41. Campaign poster for “Romanians are smart” campaign, taken from Marc van Gurp “Romanians are Smart”.

Google regarded such actions as infringements and disabled Autocomplete suggestions for the query [Romanians are]. Google’s removals are a reaction to bad publicity; however, the status of Autocomplete suggestions becomes more complex when viewed in legal terms.

The legal precedents regarding Autocomplete are summarised by Stavroula Karapapa and Maurizio Borghi in their 2015 review of court cases regarding Autocomplete, “Search Engine Liability of Autocomplete Suggestions: Personality, Privacy and the Power of Algorithm”. Their article reviews “over 25 cases that have been brought before courts in more than 10 jurisdictions. Four of these cases have

reached the relevant Supreme Court (one in Germany and three in France)" (261) in which cases have been brought against Google specifically regarding Autocomplete suggestions. The focus of each case varied, some suits were brought by individuals with unflattering or criminal suggestions attached to their name (for example, bankruptcy, fraud or rape), others were brought by companies or lobby groups regarding Autocomplete suggestions that promoted certain behaviour, for example the "French phonographic industry lobby (SNEP) […] started proceedings against Google for providing the suggestions 'Megaupload', 'Rapidshare' and 'Torrent' [legal services that can be used to illegally distribute music] alongside names of performers and titles of songs" (286). The cases addressed Google's liability, freedom of information, freedom of speech, public right to know, and present a situation in which almost every case has reached a different conclusion, even for those within the same countries. As Karapapa and Borghi describe:

> Besides the fact that courts may reach different conclusions as to the liability of the search engine, the *logic* they follow in reaching a judicial outcome varies substantially to an extent that no solid judicial trend can be assumed. This is not only because laws protecting personality and corporate rights differ from jurisdiction to jurisdiction, and there are also variations on a case-by-case basis, but also because of the ambiguity as to whether algorithmically determined word combinations can actually bear a connotative meaning that could thereinafter result in legal implications. (263 emphasis mine)

The significant difference in these cases is how the outcome of an algorithm can be defined. In cases in the United States, Autocomplete suggestions have been

successfully defended as protectable under the first amendment as the “algorithm itself is written by humans and therefore embodies a human editorial judgement” (267). Whereas the opposite perspective was taken twice by the French Supreme Court, dismissing cases regarding defamation because it was deemed that the “word combinations generated by the Autocomplete function are not an expression of thought or intellectual reasoning, but a technical method to facilitate the search” (278). There is no agreed upon definition of the kind of object Autocomplete suggestions are; such results are a complex interplay between automated technical systems, human-directed editorial engineering control, and the aggregation of vast numbers of linguistic texts: previous search queries and web content, both born-digital text and digitised language. It should also be noted that all these cases predate the moment (between February and June 2016) that RankBrain, Google’s machine learning neural network moved from its position triggering on 15% of queries that were unique to become the third most important signal that triggers on every search. The legal proceedings have a difficult enough time deciding how to define an algorithm, let alone how the outcomes of neural network should be considered and where the responsibility of those outcomes should lie.

The most recent significant distinction made was the outcome of a case in 2014 between the Spanish Data Protection Agency (AEPD) and Google Spain in the Court of Justice of the European Union (CJEU). Rather than addressing how algorithmic information might be defined, the case focused specifically on the rights of individuals and of the wider public interest. The case found that each example should be viewed on a case-by-case basis and set the precedent for Google’s inclusion of a “report inappropriate predictions” link, mentioned above, that allowed individual users to report specific suggestions that they considered inappropriate. In

addition, the court extended the “ground to request for removal to instances where Autocomplete suggestions do not need to convey a defamatory meaning or false factual information” (283). It was from this case that the now infamous phrase *right to be forgotten* became a point of academic, legal and public debate. The ability to request the removal of suggestions has not, however, led to a situation that necessarily favours individuals. Karapapa and Borghi (286) describe cases in which Google’s refusal to remove certain suggestions has been upheld on the basis of public interest. The situation is complex and as mentioned earlier the direction of Google’s information processing, via neural networks, will only continue to add further complexity.

These associated issues demonstrate why the earlier examples of specific female scientists, whose names Autocomplete suggested terms such as [husband], [wife] and [wedding], represent a particularly important example. The right for removal does not require that the listings present false information or defamatory meanings, which should allow such measures to cover stereotyping. However, removals are considered on a case-by-case basis; many people claimed via Twitter to have submitted removal requests to Google to no avail. Although, even if Google did redact these suggestions, on a case-by-case basis, the suggestions without requests would continue to appear and become even more noticeable. The individual women affected are unlikely to take Google to court over their suggestions, considering the cost, risks, and potential publicity involved. The findings of Débarre outlined above, which show that the names of specific female scientists more commonly result in Autocomplete suggestions of “husband”, “wife”, “wedding”, and “married” than their male colleagues, are of a different kind to the searches used in the social campaigns of the UN Women, or legal challenges of the SNEP. Débarre’s

study demonstrates that Google's Autocomplete perpetuates particular attitudes towards women that can only be viewed in aggregate. These examples may be more indirect than the visible misogyny of the suggestions for [women should…], but their impact may be more insidious. Such examples highlight an important issue regarding who should have access to and/or control information regarding individuals, particularly when this information might not be specifically about those individuals, but instead, represent wider cultural attitudes or beliefs. This issue will be taken up, in the context of relevance, at the end of this chapter and more fully in chapter four. For now, this chapter will continue its focus on Autocomplete, now addressing the importance of speed.

5.0 Speed

In 2009, a year after it had been made a permanent feature of Google Search, engineers made a number of significant changes to the Autocomplete tool, which at that point in time was still named Google Suggest. These were announced to the public with one main reason given: speed. As with the sentimental wedding music anecdote used to introduce users to Autocomplete, these new 2009 features were introduced by the product managers Effrat et al. using a similar focus, titled: “Faster is Better on Google Suggest”. The idea that faster is always inherently better is not only a common cliché but one which enables measurable assessment. Other common technology truisms such as easier is better (user friendly) and beautiful is better (design principles) cannot be *proved* in the same manner that speed can. Therefore, focusing on speed narrows the scope of assessment; Google’s blog post begins:

> As we prepared to write this post, we discovered a common childhood passion for fast things: high-speed trains, roller coasters, firetrucks, and more. That may be a key part of why we're so excited to be working on Google Suggest, since it saves time by giving suggestions as we're typing our searches. (Effrat et al.)

The false naivety of this opening links their changes not only to the prelapsarian state of childhood but also configures time, and its compression, in terms of efficiency (trains), excitement (roller coasters) and safety (firetrucks). Google’s post regarding the changes focuses on time-saving as an objective measurement of

improvement, rather than appeals to quality, which allows them to sidestep the subjective issue of what kinds of behaviours Autocomplete is designed to elicit. However, time acted as a red herring: the major changes made to Autocomplete in 2009 significantly changed the functional role of Autocomplete in a more complex way, making it more personalised, localised and situational. The impacts of such changes will be addressed in the following section on relevance, as well as in chapter four. Before moving on to those issues, I will spend a little more time discussing the role of time, as one of the only academic explorations of Autocomplete, Hillis et al., argue for a strong connection between Google's definition of relevance and the company's focus on speed.

5.1 Speed and Judgment

One of the main trajectories of Autocomplete is its role in speeding up searches. From early descriptions that users would spend less time typing queries, to later descriptions that Autocomplete suggestions would make users' queries more specific and thus shorten the entire information retrieval process. The two most significant time-related milestones were the implementation of Google Instant (functional between 8/9/2010 – 26/7/2017) and Google Now (established 9/7/2012 and rebranded as 'the feed' on 6/12/2016). These two technologies provide an insight not only into the centrality of speed and its relation to relevance, but also to the shifting nature of Search more generally, as it moves from a desktop situated activity to a mobile one. Such a shift is important in order to understand the rise in contextualising and personalising results described in the following chapter.

Google Instant was a direct evolution of Autocomplete and its accompanying logic. Released in 2010, Google Instant function was a feature that began searching and presenting results before a user had completed their query, using the top Autocomplete suggestion as its full presumed query. On its release, Google Instant was made the default mode of accessing Google Search for all users. At the press conference announcement detailing its release, Marissa Mayer, then Google's vice president, focused on speed and on the way that search time could be shaved down by minute increments. Mayer detailed the average speeds of each part of searching for a query (see fig. 42 for her accompanying slide) and took issue with the average time, nine seconds, users took to type a query, underlining her point with the nonsensical fragment: "never underestimate fast" (13:03). Mayer went on to argue that through Google Instant, by combining the total time saved globally, "you actually will be able to save eleven hours for every passing second" (22:10), the phrasing of which was directed not at the aggregate but to an individual "you" and your "passing" seconds. What the global community might choose to do with this time was not addressed.

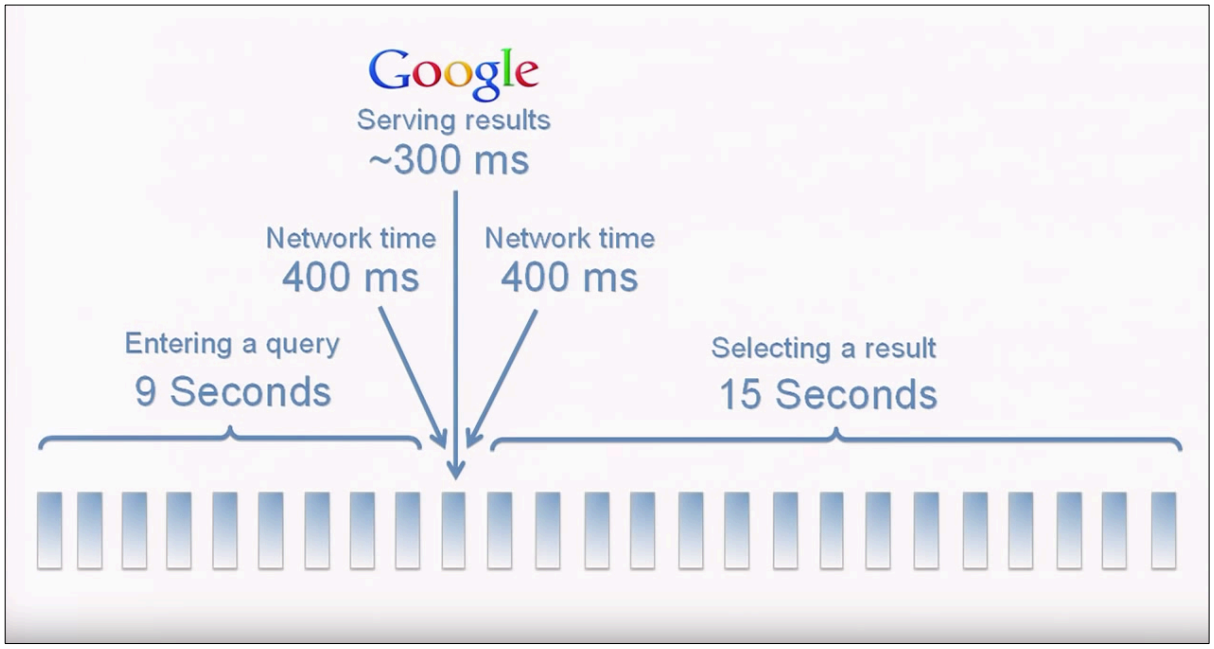


Figure 42. Presentation slide at 13:03 from Marissa Mayer's "Google Instant Launch Event" detailing the average time taken for the three portions of searching for a query.

Google Instant relied on Google Autocomplete's suggestions being accurate for the majority of users in order for the automatic search to be considered a benefit rather than a distraction. Mayer addressed the confidence Google had in the Autocomplete suggestions:

> a lot of people think Google Instant is search-as-you-type, but it's actually search-*before*-you-type […] we're actually predicting what query you're likely to do and giving you results for that […] there's even a psychic element of it, in that, we can actually predict what you're likely to type and bring you those results, in real time. (18:33)

The argument was not made that Google Instant would change or direct the queries of users, but that the aggregation of information held in Autocomplete's suggestions would be powerful enough to produce a "psychic" meshing between user and interface. It is this kind of speed-consequential effect that Hillis et al. focus on in their critique of Autocomplete and Google Instant, a critique that also informs this thesis' later discussion regarding relevance. In particular, Hillis et al. draw on the work of Scott Lash, arguing that Google's attempts to speed up the interactions between users and the search engine constitute the conditions for play. As Lash argues in *Critique of Information*, "To play is to be *so* interested, *so* involved immediately as to rule out the possibility of judgement. Judgement involves always a separate and neutral instance. It presupposes a culture of representation. Play … does not involve this" (160). Play, a form that Lash usually considers with ambivalence, for Hillis at al. takes on a decidedly negative role. Such a characterisation fits Mayer's position, as she gestured towards her slide (pictured as fig. 42) arguing that the total time for a search, around 24 seconds, is too slow and is time that must be saved. Hillis et al. argue that when immersed in activity of play,

> She or he cannot generate the objective, reflective, reflexive distanced judgement of the transcendental ego variously set forth by Kant, Hegel, and Husserl. Google's model of relevance does not support an epistemological position from whence one might observe reflexively in order to make aesthetic or critical judgements. (74)

Although I am wary of such an absolutist appeal to unmediated objective attention – this thesis' previous chapters aimed to demonstrate that attention and thought are

always enmeshed, both socially and technologically – reflexivity and critical judgements are evidently the victims of Mayer's "psychic" "search-*before*-you-type" (18:33). The heterogeneous vantage point that Autocomplete *could* offer is side-lined; although still visible, other suggestions are dominated by the top suggestion, the result of which fills the entire screen. Although this notion of play is partly useful, both Mayer and Hillis et al. seem to have a rather singular vision of who the average user of Google might be, and the expertise required to rely on Google Instant in the way originally intended.

Google Instant had its detractors; in an article covering its discontinuation, titled "RIP Google Instant Search, You Were Never Necessary", journalist Rhett Jones referred directly to Mayer's original claims regarding Google Instant's psychic abilities, arguing that, "Yes, it had a psychic element, in the way that someone who constantly interrupts you thinking they know what you're going to say next has a psychic element". Although Google did not cite a lack of popularity as their reasoning for discontinuing Google Instant, we can never know what impact the potentially disruptive loading and reloading that caused the whole screen to flash and stutter had on different users. For this reason, I argue that play might be a useful model for how it altered the search experience for *some* users, but that this might be limited to specific demographics with particular technical familiarity. Assigning play central conceptual importance, in regard to Autocomplete, prioritises speed and high levels of expertise. In doing so, play emphasises the capacity for Autocomplete to assist searchers by accelerating, rather than changing, their search queries. As discussed in chapter one, it is impossible to fully know the intentions of users; therefore, it is hard to classify whether following an Autocomplete suggestion represents a deviation from a user's original intended query or not. Consequently, this tension,

between acceleration and change, is at the heart of every Autocomplete suggestion. However, this is only the case if the Autocomplete suggestions align with the attitude of a user. Autocomplete is designed to function unobtrusively but, as has been demonstrated throughout this chapter, Autocomplete becomes an object of attention in the moments when the tool fails to match the expectations of users. Martin Heidegger's tool analysis provides a conceptual model for theorising these moments of failure.

Heidegger's tool analysis can be used to highlight how individuals, or Dasein to use Heideggerian terminology, encounter and engage with the world around them. For Heidegger, when technologies fail their users the attentiveness toward that object changes from a usage that attends to them fairly transparently to focusing directly or interrogatively on that object, a shift Heidegger describes as a movement from ready-to-hand to present-at-hand. As Graham Harman outlines:

> For the most part, Dasein encounters entities that are not present-at-hand, but ready-to-hand. Dasein does not usually stare at things or analyze them theoretically, but uses them and takes them for granted. In any given moment, most of us are not thinking about the chair we are sitting in, the floor that supports it, the solid earth that keeps us alive. Instead, we take these things for granted and focus our attention elsewhere […]
>
> equipment usually hides from us. It is inconspicuous or unobtrusive. Usually, only bad equipment makes us notice it frequently, such as when ceilings are too low and we bump our heads too often. But equipment also malfunctions sometimes. Cars break down; hammers fall apart or wine glasses shatter; bodily organs suddenly fail us. It is mostly in these moments

> that equipment first becomes conspicuous and draws our attention to it. (62-63)

When the Autocomplete suggestions for a query match the outlook of a user they can be easily ignored as unremarkable. By extension, when Google Instant's page retrievals successfully predicted a user's query its action can be described as ready-to-hand, almost invisible to a user, as an extension of their intentions. However, any Autocomplete suggestion that differs from a user's intended query acts as a moment of failure: Autocomplete has not *broken* in the sense of a shattering wine glass, but instead provides a moment of rupture. When intention or outlook fall out of synch with the Google Autocomplete suggestions presented, those suggestions become the object of a user's attention: present-at-hand. Google Instant enhances such a rupture by filling the user's whole screen with the result that they consider unrelated to their search; the less innocuous the suggested query the more present-at-hand the tool becomes. Paul Baker and Amanda Potts explore the different kinds of stereotypes that emerge in Autocomplete suggestions. A pertinent example for this current discussion is the scenario in which,

> if someone wanted to use the Internet to find the answer to the question 'Why do black holes exist' when they start typing the question into Google, after the first three words, they would be presented with a number of auto-complete suggestions including 'Why do black people have big lips' and 'Why do black people like chicken'. (201)

Their observation was from 2013, the same search in 2017 leads to similar suggestions (see fig. 43).

Figure 43. Autocomplete suggestions for [why do bla] accessed 15/8/2017.

Baker and Potts do not reference Google Instant, they are interested in the aggregation of stereotypes rather than the phenomenological aspects of the search experience on an individual basis. However, in the context of Google Instant, such a result being the first suggestion has significant consequences. For the seven-year period in which Google Instant was the default setting of Google Search, a user typing their search for [why do black holes exist] would be looking at a full page of results for [why do black people like fried chicken] as soon as they had entered [why do bla]. This kind of present-at-hand rupture would have been relatively common, as evidenced already by some of the examples shown. For Heidegger, attending to a tool as present-at-hand allows for a critical or reflective position that is usually unavailable when a tool is used as ready-to-hand. The popular vlogger genre

outlined at the start of this chapter is, after all, all about the shock, strangeness and humour that top Autocomplete suggestions provide. These are moments of reflection about how Google works, how language functions both on- and offline, and how distributed centres of power permit or curtail the perpetuation of various attitudes. Using the terminology of ready-to-hand and present-at-hand allows the discussion to move away from the rigid dichotomy of success and failure, terms that present the situation from Google's point of view. The default nature of Google Instant represented a judgement that Autocomplete suggestions were, most of the time, a successful match for users, or ready-to-hand. The occasions in which Google Instant provided present-at-hand results are not simply moments of user dissatisfaction but instead opportunities for critique and reflection about the whole of Google's discursive regime. This consequence is a likely part of why Google Instant was discontinued in 2017, although this line of reasoning is unlikely to be included in any official press statements.

Google addressed the reasoning behind discontinuing Google Instant in a statement to the site *Search Engine Land*:

> We launched Google Instant back in 2010 with the goal to provide users with the information they need as quickly as possible, even as they typed their searches on desktop devices. Since then, many more of our searches happen on mobile, with very different input and interaction and screen constraints. With this in mind, we have decided to remove Google Instant, so we can focus on ways to make Search even faster and more fluid on all devices. (qtd. in Schwartz "Google has dropped Google Instant Search").

In the *Search Engine Land* article that contained Google's statement, Barry Schwartz referred to Google's reports, mentioned already in this thesis, that the majority of Google searches now take place on mobile and tablet devices. As journalist Nick Statt described, given the size of mobile screens "Instant Search doesn't make as much sense given we use our fingers and virtual buttons to interact with software, and trying to load a results page on top of the onscreen keyboard isn't exactly good user experience design". In addition to hardware constraints, Google's decision was based on how users search using mobile devices. This leads to the second descendant of Autocomplete predictions, Google Now.

Google Now (rebranded as 'the feed' from 6/12/2016) is both an extension of the logic of Autocomplete and its "psychic" aspects as well as a spiritual successor to the kinds of changes made to Autocomplete in 2009 that will be discussed below. Google Now was designed as a response to the increasing percentage of searches that are carried out on mobiles. Google Now and 'the feed' represent the step beyond Mayer's "search-*before*-you-type" using information to provide users a curated list of news or articles that reflect their interests, but also contextual information based on location, time and other metrics, such as automatically showing the nearest restaurants to an individual who is out of their home around a meal time or the upcoming buses around the time a user usually commutes to work. The release statement regarding Google Now rebranding as 'the feed' charts a course between the Scylla and Charybdis of irrelevance and dystopia:

> People have long turned to Google to get answers, learn about the world, and dig deeper on topics they're passionate about. Today, we are announcing a new feed experience in the Google app, making it easier than ever to

> discover, explore and stay connected to what matters to you—even when you don't have a query in mind.
>
> Since introducing the feed in December, we've advanced our machine learning algorithms to better anticipate what's interesting and important to you. You'll see cards with things like sports highlights, top news, engaging videos, new music, stories to read and more. And now, your feed will not only be based on your interactions with Google, but also factor in what's trending in your area and around the world. The more you use Google, the better your feed will be. […] As the world and your interests change, your feed will continue to grow and evolve along with you. (Thakur)

Such a statement represents an overall shift to personalising information, using contextual signals to deliver information to individuals that affects them. Using machine learning to "anticipate what's interesting and important to you" carries significant philosophical and social weight. It represents a fine-tuned version of Autocomplete as it originally existed, using data for predicting queries; the approach of 'the feed' is that predictions can be more precise and potentially more accurate if they are tailored to specific individuals. Whereas Autocomplete's strange present-at-hand suggestions might provide a vantage-point for reflection in highlighting some wider linguistic arrangement, the logic of 'the feed' can use machine learning to tighten the informational loop between users and their devices in order to keep such access ready-to-hand in a way that feels neutral or comfortable. The link between this logic, Autocomplete suggestions, and one of Google's wider drives as a company is the notion of relevance, the significance of which is discussed below.

6.0 Relevance for Autocomplete

The changes that were made to Autocomplete in 2009 focus on making Autocompleted suggestions less homogenous, responding more directly to the context of the user. Chapter four will address relevance in the context of Google Search as a whole; therefore, although this section will focus directly on the relationship between Autocomplete and relevance, in order to bring this chapter's exploration of Autocomplete up to the present, it also acts as a tributary into some of the wider issues discussed in the following chapter.

In addition to time-saving, Effrat et al. outline four other changes made in 2009: "suggestions on the results page", "personalized suggestions", "navigational suggestions", and "sponsored links in suggestions". These alterations fundamentally changed the function of Autocomplete, away from the generalised list of suggestions and towards a contextual framework that relates to the general shifts of Google's search engine from then onwards. These shifts were further enhanced by changes made in response to the growing use of mobile and declining use of desktop devices, with which to search queries.

Beginning with "suggestions on the results page", this meant that the suggestions provided in Autocomplete corresponded to the page the user might be on at the time. Their example is while searching Google from a Google Search results page concerning roller coasters (notice how their earlier example reinforcing speed and fun recurs) entering [b] into the search field would provide the user with suggestions concerning rollercoasters, rather than just the regular suggestions for the letter b (see fig. 44).

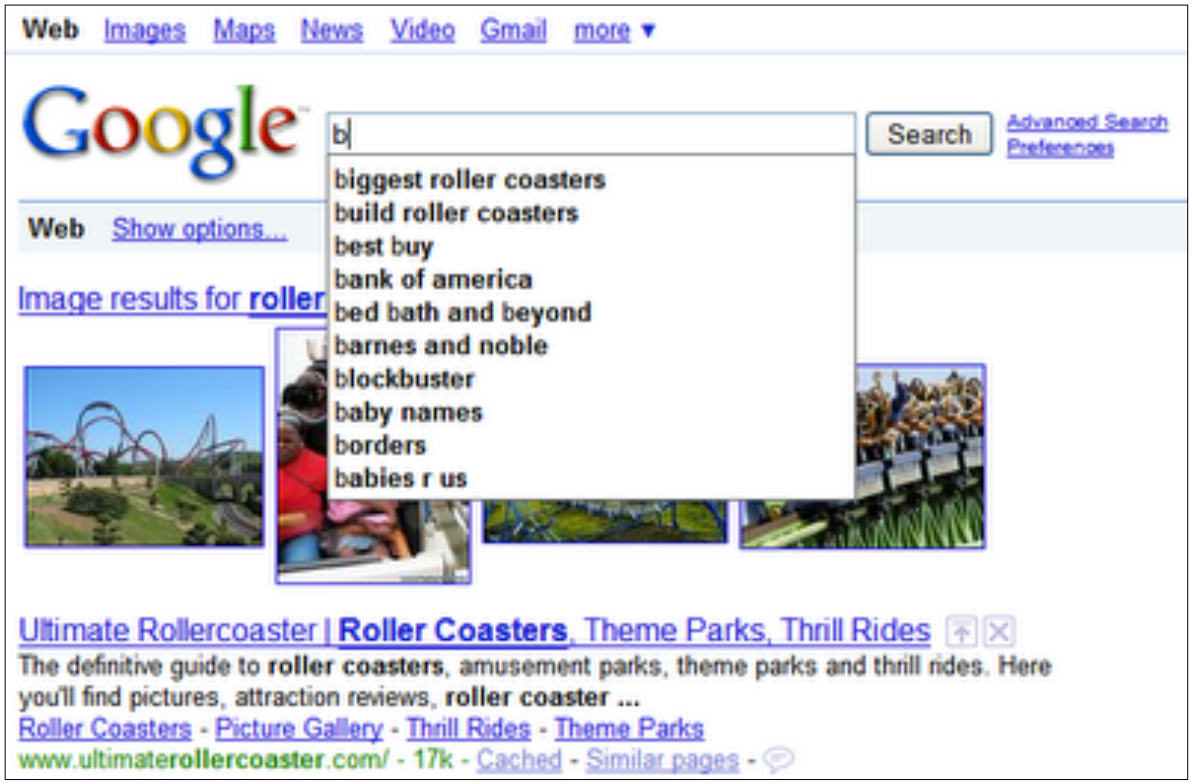


Figure 44. Autocomplete results for [b] while on a Google results page for rollercoasters, taken from Effrat et al. accessed on 05/04/2014.

This provides a personal narrative structure to searching that did not exist previously. Whereas Autocomplete pre-2009 provided a universal list of suggestions, this feature tailored suggestions to the user to act as a directional aid, connecting searches and limiting the user's field of view. This relates to change number two, referred to as "personalized searches" (see fig. 45), and promotes repeat search behaviours.

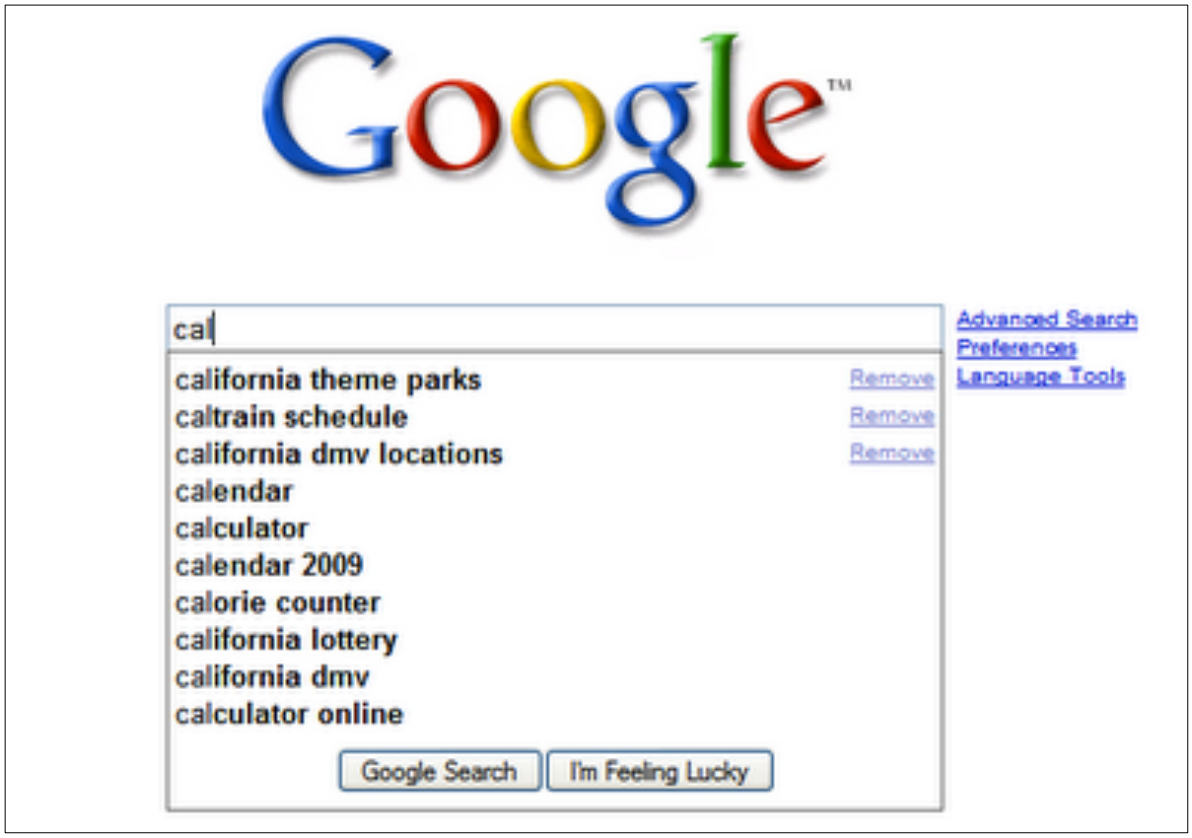


Figure 45. Demonstration of personalized search, taken from Effrat et al.

This tool does not differentiate between search queries from the past which were helpful or unhelpful for the user, a difficult quality to measure. Neither does this feature base the reminder of previous searches around topic or theme, instead, it simply follows the letters or words beginning a search query. In some situations, it might be argued that this addition makes searching faster, in particular when the user is aiming for a specific page they have visited before or visit regularly. It is less useful for searches which require repeated queries over a period of time, news for instance, where a more accurate search which would outline a developing situation, for example, a search for [Ukraine] might have been useful in the early stages of the 2014 Ukrainian revolution, to get an overview of a number of websites pertaining to the country. However, to keep up with the revolution as it unfolded varying searches

such as [Ukraine Russian intervention], [Ukraine Crimea], and [Ukraine US reaction] would be more useful than a repeated search of [Ukraine].

The last two main feature additions of 2009 relate to the conceptual merging of search and web address functions. “Navigational suggestions” enabled specific websites to be included in Autocomplete suggestions meaning that, for some websites, users could travel directly to suggested sites without even searching. Finally, the inclusion of “sponsored links in suggestions”, a feature which has now been removed, allowed paid advertising to be included in the drop-down list of suggested terms. The exact reasons for why this particular feature was decommissioned are unknown but decisions were made that we might assume have to do with the perceived neutrality of Google Search and the perception of Autocomplete as a functional part of Google Search rather than as a moneymaking venture.

An additional change made in 2009 represents the further direction, discussed above, regarding localisations. Originally Autocomplete was simply divided by language, showing the same suggestions wherever the user was based. Google changed Autocomplete in 2009 to reflect localisations, for example “If you type [liver] in the U.K., you're probably a Liverpool fan (but in the U.S. you'll get more suggestions about liver diseases)” (Kadouch). This logic of aiming towards more specific and culturally situated results reflects on the future iterations of Autocomplete and particularly its link to Google Now and ‘the feed’. The phrasing of the above quotation, stressing probability, contextualises such a decision as *functional* rather than outwardly editorial in regard to wider epistemological and cultural questions. In this way, such a logic fits within the Heideggerian framework of technological interaction that aims towards fostering ready-to-hand experiences.

These issues fit within Google's wider criterion of relevance and serve to underpin larger aims that go beyond the specific instance of the Autocomplete tool. These wider issues will now be developed in the context of search results in the following chapter.

Conclusion

This chapter began by demonstrating the imbalance between public and academic interest in Google's Autocomplete tool and focused on redressing this disparity. The initial section challenged the claim, perpetuated by vloggers, games such as Google Feud and idiots.win, and other examples, such as *Wired*'s Autocomplete Interviews, that Autocomplete suggestions ngprovide an insight into a shared public consciousness. There is much evidence to dispute the claim that Autocomplete's suggestions are an indication of query volume – the number of times a specific query had previously been searched; however, the idea that such suggestions tell us something about language use, topic correlation and public interest may still hold, although not for the reasons currently circulated. This chapter asserts that Google's general drive towards neural network machine learning and wide-scale data collection, including their Google Books project, is an important factor in the kinds of language, topics and attitudes that are suggested by Autocomplete. The significance of RankBrain raises many centrally important questions regarding liability, responsibility, and the perpetuation of stereotypes. Autocomplete suggestions act as an insight into the ways in which influential technology companies are aggregating language in a way that has serious social consequences. The default nature of Autocomplete, the lack of transparency regarding its criteria, and the absence of any way for users to alter the results given (other than reporting specific instances) speaks to the power imbalance on the web, in which a single company has control over how crucial issues such as race, sex and gender are contextualised. In addition, if Google's Autocomplete suggestions *are* created by machine learning rather than algorithms then even Google's own engineers may not be able to properly outline the

computational reasoning behind certain suggestions. Finally, Google's recent modifications to Autocomplete demonstrate the expansion of their criterion of relevance, as a way of making search personalised and contextual. The implications of such a direction are important for the wider discourse of online citizenship, the resulting issues of which are addressed directly in the following chapter.

Chapter Four:

Search Engine Results and the Plurality of Googles

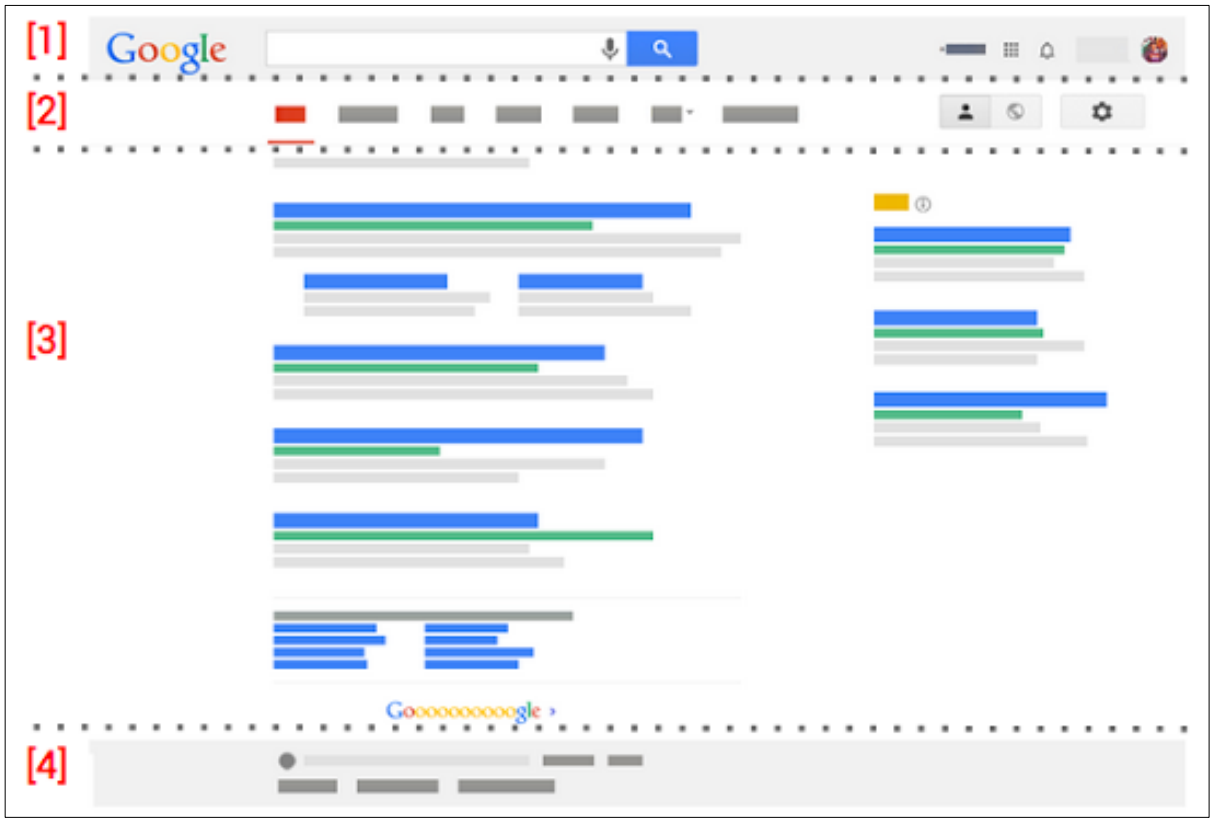


Figure 1. Diagram of a search results page, taken from Google “Google Search Help: Google Search Results Page”. Accessed 10/02/2017.

Introduction

This chapter addresses search engine results and reviews the attitudes towards a particular question that recurs, in different guises, throughout various areas of search engine research. In reference to the above image (fig. 1), the question in its broadest

terms is: *what should appear in section three of the diagram?* Google's list of search results is, of course, the obvious answer; however, an attempt to be more specific leads to an intellectual garden of forking paths. The question could equally be framed, *what kinds of results should search engines provide?* Or even, *what are search engines for?* Such questions implicitly underpin much current research in a range of fields; however, the attitudes that substantiate these perspectives are rarely the focus of existing literature. Such attitudes direct disciplinary approaches to research questions, as well the corresponding methodologies. For example, researchers from the discipline of information retrieval measure the "performance" of search engine results numerically by "making judgement[s] about the value, importance and quality" (Goel and Yadav 7) using statistical models. Many legal theorists judge search engine results in terms of whether they are "fair and diverse" (Diaz 15) and without clear economic biases that relate to competition laws. Various cultural theorists argue that search engines should "give voice to diverse social, economic, and cultural groups, to members of society not frequently heard in the public sphere" (Introna and Nissenbaum 169) and rate search engine providers and specific results accordingly. Software developers, outside as well as inside Google, have a range of user-focused metrics that dictate whether or not new algorithms or ranking systems offer "better performance […] and can improve on Web search" (Teevan et al. 449) through measurements of click-rates and post-experiment interviews. This is to say nothing of the large media interest and wealth of journalism that regularly reports on what search engines should and should not be providing for queries.[1] Because search engines are used for such a wide range of purposes that

[1] An example that frequently recurs in the news cycle is whether or not sites that promote holocaust denial should be listed as Google results. For an example see Frank Pasquale's article "From Holocaust Denial to Hitler Admiration, Google's Algorithm Is Dangerous", in which he lists five steps that Google and Facebook "should take to move toward more transparency and accountability", a

cover various subject areas, researchers, journalists, and the public often talk at crossed-purposes as a result of assumptions that are made about what search engines are for and what constitutes a good search result. In addition, there are a number of methodological challenges that affect different disciplinary approaches in various ways, as well as certain evaluative norms that vary between journalists, legal theorists, computer scientists and between academic disciplines. For this reason, this chapter focuses as much on how researchers can study and evaluate search engine results, as on the results themselves.

The chapter outline is as follows: first, the chapter addresses some of the key methodological challenges that develop from studying search engines. Second, it addresses the evaluative frameworks that are used to judge search engine results. Chapter one of this thesis outlined the wide range of functions that search engines are used for, by drawing from research that subcategorises query types, such as the 2008 study by Jansen et al. “Determining the Informational, Navigational, and Transactional Intent of Web Queries” (see fig. 2).

---

statement which demonstrates a number of implicitly held beliefs regarding how search quality ought to be defined.

Definitions of classifications of Web queries

Levels	Examples of queries
Level one	
• (I) Informational: queries meant to obtain data or information in order to address an information need, desire, or curiosity • (N) Navigational: queries looking for a specific URL • (T) Transactional: queries looking for resources that require another step to be useful	• Child labor law • Capitalone • Buy table clocks
Level two	
• (I, D) Directed: specific question • (I, U) Undirected: tell me everything about a topic • (I, L) List: list of candidates • (I, F) Find: locate where some real world service or product can be obtained • (I, A) Advice: advice, ideas, suggestions, instructions • (N, T) Navigation to transactional: the URL the user wants is a transactional site • (N, I) Navigation to informational: the URL the user wants is an informational site • (T, O) Obtain: obtain a specific resource or object • (T, D) Download: find a file to download • (T, R) Results page: obtain a resource that one can printed, save, or read from the search engine results page • (T, I) Interact: interact with program/resource on another Website	• Registering domain name • Singers in the 1980s • Things to do in hollywood ca • PVC suit for overweight men • What to serve with roast pork tenderloin • match.com • yahoo.com • Music lyrics • mp3 downloads • (The user enters a query with the expectation that 'answer' will be on the search engine results page and not require browsing to another Website) • Buy table clock
Level three	
• (I,D, C) Closed: deals with one topic; question with one, unambiguous answer • (I,D, O) Open: deals with two or more topics • (T, O, O) Online: the resource will be obtained online • (T, O, F) Off-line: the resource will be obtained off-line and may require additional actions by the user • (T, D, F) Free: the downloadable file is free • (T, D, N) Not free: the downloadable file is not necessarily free • (T, R, L) Links: the resources appears in the title, summary, or URL of one or more of the results on the search engine results page • (T, R, O) Other: the resources does not appear one of the results but somewhere else on the search engine results page	• Nine supreme court justices • The excretory system of arachnids • Airline seat map • Full metal alchemist wallpapers • Free online games • Family guy episode download • (As an example, a user enters the title of a conference paper in order to locate the page numbers, which usually appear in one or more of the results) • (As an example, a user enters a query term to check for spelling with no interest in the results listing)

Figure 2. Table taken from Jansen et al. "Determining the Informational" demonstrating their three-tier classification system.

This chapter's second section seeks to continue that earlier discussion by drawing on a range of critical perspectives from various disciplines, in order to provide examples of the implicit positions adopted in regard to what search engines *should* be. In doing so, I argue that attitudes towards search results fit into one of two mutually exclusive perspectives, which I describe as Group A and Group B thinkers. The first, Group A thinkers, argue that search engines are fundamentally democratic due to their underlying technological construction and, therefore, should consequently aim towards enhancing that dimension; some even take the more

extreme stance that the aggregated information of the web, if complete enough, would represent some form of objectivity, not unlike Plato's Idealism discussed in chapter two. The thinkers that I categorise as belonging to Group B describe the search engine paradigm as one that provides relativistic truths, which require personalisation and the alignment of results on a contextual and individual basis. These attitudes are outlined with particular reference to the criterion of relevance, in order to fully develop the discussion that was established at the end of chapter three. In addition, Group B thinkers replicate certain attitudes that align them with an Aristotelian epistemology; therefore, such a contrast speaks back to the long history of memory technologies outlined in chapter two. However, it is important to note that these contemporary attitudes develop from specific contexts, the discipline of information retrieval (IR) for example; therefore, this later analysis is wary of pushing on this parallel too firmly, for fear of warping Group A and B perspectives into ahistorical generalisations.

After outlining these positions, the third section of this chapter argues that these attitudes have coalesced around particular disciplines and specific methodologies, the employment of which has shaped the academic conclusions regarding search engines. There is a lack of consistency between the ways that various disciplines have set out to evaluate search engine results. Therefore, the different modes of measurement and analytic frameworks used have produced a range of incomparable results leading to an increased distance between disciplinary perspectives. The final section of this chapter aims toward remedying this situation. Currently, there is no academic work that surveys a range of search engine results, while altering certain metrics, such as location, in order to describe and analyse the variation. The small number of studies that have been conducted set out their ideal

version of what a set of search engine results *should* be (diverse, relevant, accurate, or any other highly variable criterion), collect a range of results and conclude whether or not they succeed in reflecting their initial expectations. In particular, although it is generally agreed upon by academics and journalists that search engine results are different for each individual, depending on their context, there have been no attempts to provide a rigorous, comparative analysis. There is a lack of academic work that seeks to describe the variations between search contexts, comparing, for example, the search results served to a French speaker searching Google in Morocco to an English speaker searching Google for the same query in France. Therefore, the final section of this chapter comprises a study conducted to highlight search result differences and provide an open-ended analysis of the kinds of webpages served as Google search results. Although there are many challenges to considering algorithmically produced results as representative of more general trends, as will be discussed below, the results collected aim to raise questions regarding Google's role and function. Specifically, my study, conducted 24 November 2015, collected search results for particular queries while changing certain criteria: the location that the query was sent from, the language used, and the kinds of phrasing deployed. The range of differences found between contexts demonstrates how search results can differ, from person to person and place to place, and identifies which factors produced the most significant differences. This chapter concludes by comparing these results with the results of the same searches conducted 24 months later, on 22 November 2017. The comparison demonstrates a number of changes and is accompanied by a speculative analysis concerning the reasons for such a shift. In particular, this section draws from the conclusions of

chapter three to hypothesise that Google's overall search results may have changed due to their increasing reliance on machine learning technology.

1.0 Methodological Challenges

Researching search engine results carries a number of methodological challenges. These challenges create inherent problems regarding how researchers can measure or assess web search engine results, which further obfuscates a general understanding and prevents the emergence of shared interdisciplinary standards. In addition to the different disciplinary stances outlined above, this chapter will engage with the challenges that influence and restrict the kinds of studies that can be made of search engines. Eszter Hargittai, in her introduction of the 2007 special issue of the *Journal of Computer-Mediated Communication*, outlines a number of the methodological challenges central to studying search engines, all of which are still prominent today. First, Hargittai highlights that "data tend to be proprietary and are rarely released to researchers" ("The Social, Political" 771) and refers to the 2006 AOL query log release, mentioned in chapter one, that led to the firing of employees, a class action lawsuit and heightened precaution within the industry. In many cases, analysing why a certain query provides the set of results that it does requires highly speculative reasoning. Second, Hargittai continues, there is no independent "comprehensive listing of all existing sites" (772). Commonly used methods in IR rate algorithms or ranking systems by comparing the retrieved results with a wider corpus or dataset. Therefore, when judging the quality of a set of search engine results, users and researchers can only compare results to other results *also* found using a search engine, mostly a rephrased query submitted to the same engine or an alternative search engine. As noted earlier, the size of the deep web, those pages unfindable by web crawling spiders, is unknown. Also unknown are the kinds of pages that might be indexed by a search engine and returned as results but never

prominently enough to be found. See fig. 3 for example, which demonstrates the banality of the web's embarrassment of riches. Given the statistics, noted earlier, that most users follow the top result and barely any even navigate to page two, anything could be hidden in plain sight, as one of the other 123,999,990 results.

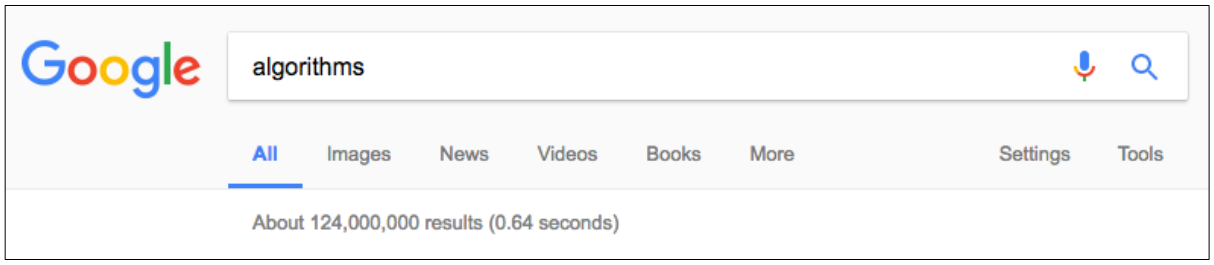


Figure 3. A search for [algorithms] on 22/08/2017 returned "about 124,000,000 results". Search performed 22/08/2017.

Again, "because the algorithm of search engines is proprietary information, it is impossible to know what gets covered and what does not" (772); organisations such as the Internet Archive, founded in 1996, represent attempts to create a historical record of the web. However, unearthing webpages that might be indexed by Google but never shown prominently enough in their results, relies on search engine technologies that require either access to Google's technical information or comparable resources, neither of which a small non-profit organisation would have access to. Third, Hargittai notes, search engines are "moving targets" (772). Not only does the web grow, but Google makes around "500-600" changes to their algorithms a year (Moz). The Internet Archive's Way Back Machine can archive snapshots of pages throughout time (see fig. 4), but there is no way to return to historical Google search states. If a researcher claims a particular search resulted in a set of results, there is no way to check or repeat their search.

Figure 4. The Internet Archive's Way Back Machine, showing a snapshot of google.com, as of 01/06/2001. Accessed 22/08/2017.

This brings us to Hargittai's final point, that

> in some cases, results change by user and user location, so a study conducted on one machine in one location by a particular user may not be possible to replicate on another machine under different circumstances, even soon after the initial query. This poses significant challenges for the replication of search results, which is a basic tenet of scientific investigation. ("The Social, Political" 772)

Search engine results cannot be thought of as universal or generalisable. There is a fundamental lack of knowledge regarding how the results of users might differ from one another and why this might be the case. Throughout this chapter, the challenges that derive from such unknowability will be addressed, as even acknowledging them

and attempting to ascertain their degree of influence provides a clearer understanding of search engines.

The second section of this chapter follows an experiment carried out on 24 November 2015 and repeated 22 August 2017 searching various queries in different languages, locations and phrasings, designed to demonstrate the differences in search results in different contexts. The findings are used to reflect on the debates of this chapter's first section, while acknowledging the shortcomings of such a method, and interrogate the complexity of judging such results. In addition to outlining statistical differences, this section promotes an open-ended analysis of such results, in light of the literature review of assessment methods, as this approach is rarely conducted. The findings correlate with other reports that, for example, altering the location of a query changes the results. Although such results are not generalisable, my study shows that some changes, for example searching in a different language, have a much greater influence on results than, for instance, searching in the same language but from a different location. In addition, the evidence collected also demonstrates a significant change in the overall search results between 2015 and 2017. In 2015, altering variables had a greater effect on results and a wide range of different pages were found. In 2017, however, the general pool of results was much smaller and changing languages or phrasing did not produce as many differences, leading to a much more standardised set of results across different contexts. I hypothesise that this change was caused by Google's shift towards machine learning neural networks (between February – June 2016) under the management of John Giannandrea, after taking over from Amit Singhal, which was discussed in the previous chapter. There are difficulties in substantiating such a claim, due to the limitations already mentioned; however, this claim presents an opportunity for

discussion. Regardless of the reason or cause of such a change, or if such a general shift has taken place at all, the examination of the interactions between local and global thought and wider cultural practice are particularly precedent, as these issues are developed further in chapter five.

2.0 What *Should* Search Engine Results Be?

Evaluating search engine results depends on particular perspectives regarding what search engines are or what they ought to be. A significant part of this develops from the disciplinary methodologies available to researchers, whereby the measuring-stick often leads the direction of study, which will be discussed in this chapter's later section. This section focuses on more explicit claims regarding the fundamental nature and function of search engines. Such claims fall roughly into two camps: the first, which argues that search engines are fundamentally either objective, neutral, or democratic and the second, which envisions "perfect search" (Battelle 252) as deeply personalised, subjective, or contextual. These two positions are dramatically incompatible; however, a range of irreconcilable attitudes are held by search engine technologists, academic researchers and journalists. These are not simply claims of preference, but rather, as the authors see it, descriptions of the inherent nature of search engine technology, as determined by their technological character. Although all the sources presented below would all agree on the technologies that underpin search – web crawling spiders, indexing and caching, algorithms such as PageRank, Markov Chains, keyword search – their conclusions regarding what a search engine *does* and how such actions are to be *judged* vary greatly. Such a discussion is important because any kind of judgement about search engine results relies on a conception of what search engines are designed to do. These two positions regarding search engines, which I refer to as group A and group B, can be roughly mapped onto the Platonic and Aristotelean positions, respectively, which were outlined in chapter two. Although the rhetoric of some technologists and engineers suggests that they would like to have their Platonic cake and eat it with Aristotelian

bites, such epistemological viewpoints are incompatible. Problems emerge due to the two groups using the same terminology but defined in incompatible ways, in particular, the terms relevance, democracy, and bias. Therefore, by unearthing the underlying argumentation, this section provides clarity regarding current search technology and problematises the teleological attitude that in the future search engines will simply be *better*.

2.1 Group A: Search Is Democratic, Relevance Can Be Measured Objectively, and Answers Can Exist Independently of Bias

The notion that Google is inherently democratic is a pervasive argument, one used by Google's marketing department as well as a range of technologists, and one that has been used to defend Google as neutral or blameless in a number of situations. Computer scientist Ed Felten's description sums up a much larger discourse. He says that, "Google is a voting scheme […] not a mysterious Oracle of Truth […] It's a form of democracy – call it Googlocracy. Web authors vote by creating hyperlinks, and Google counts the votes. If we want to understand Google we need to see democracy as Google's very nature, and not as an aberration." In the comments section of Felten's piece, he engages with various criticisms, arguing explicitly that webpages that are given a low rank "can't be Google's fault". Google used to provide the democratic metaphor as a central part of their self-description, in 2007, their "Google: Technology" page was given over to describing their PageRank algorithm in the following terms:

> PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at considerably more than the sheer volume of votes, or links a page receives; for example, it also analyzes the page that casts the vote. Votes cast by pages that are themselves 'important' weigh more heavily and help to make other pages 'important.' Using these and other factors, Google provides its views on pages' relative importance.

The workings of PageRank were outlined in chapter one, although its explanation there relied on the metaphor of academic citation, rather than votes in a democracy. Such a description, however, has gained a large degree of traction in academic discourse. One of the earliest pieces of research to address search engines within a political and social framework, Introna and Nissenbaum's "Shaping the Web: Why the Politics of Search Engines Matters", sees search engines as a way of defending the inherently "democratizing force" of the web. They describe search engine technology as a way of preventing the market from dominating the web, allowing it to "empower the traditionally disempowered, giving them access both to typically unreachable nodes of power and to previously inaccessible troves of information" (169). These early notions regarding the potential of search engines to protect the web's supposed democratic nature have spilled into wider public perceptions. Considering that much of the technical details regarding how search engines are, at best, too complex for a non-technical audience, and at worst, black-boxed as proprietary information, these perceptions have come to dominate a wider set of expectations. As Alejandro Diaz describes:

> we expect search engines to present the available information in a fair and diverse manner; we expect them, in other words, to be 'democratic.' We should ask about search engines like Google the same questions scholars have asked about traditional media: *Can underrepresented voices and diverse viewpoints be heard through the filter of search engines? What role does advertising play in the returned results? Do a few players dominate the industry?* (15, emphasis in original)

Although Diaz is ultimately critical of the broad uses of the term *democracy*, the terms of the debate have been set and their definitions expanded and conflated; in much of the discourse fairness, objectivity, democracy, and a lack of bias become muddied in with one another. Such techno-determinist notions have even been the basis for manifestos, Hiroki Azuma's *General Will 2.0: Rousseau, Freud, Google* argues that Google's monitoring, accumulation, and analysis of data provides an opportunity to institute an updating of Rousseau's concept of the general will in which "We can, and we should, update the principles of democracy to a new form by making use of the experiences of the information society" (xii). Such a perspective conceptualises Google's activities as a form of democracy that is fairer and more representative, not because the views of everyone are taken into account, rather because such a wealth of data provides insight into every individual's unconscious attitudes. For Azuma, filtering this wealth of data through a psychoanalytic framework provides an objective measure of the collective's absolute subjectivities, which enables a radical re-reading of Rousseau's ethics to be enacted and for democracy to exist in an unprecedentedly pure way:

> in the world of general will 2.0, the private, animal behaviour of the masses (database) will be aggregated and visualized through information technology and will place limits on politicians and experts' public consensus formation (deliberation). The accumulation of animal behaviour will constrain human judgement. (158-159)

Azuma's counterintuitive notion is that absolute subjectivity, if aggregated in large enough quantities and unfiltered through conventional discourse, inevitably leads to a purer and more objective description of humanity. Such a vision lies at the extreme end of such a discourse and aims to be provocative, however, its underlying logic, that with enough information and superior enough technology objectivity begins to emerge, is prevalent throughout both scholarly and public discourse.

In the scholarly context of search engine evaluation, a deterministic mode that parallels democracy has gained an overwhelming influence on current discourse: relevance. Relevance has been mentioned a number of times throughout this thesis, although its discussion has been limited. This is because the term has been used variously to mean a number of positions, that taken together contradict one another. Following on from the above democratic discussion, I will first outline how the term relevance has been deployed in a Platonic sense that represents objective relations between ideal objects. The other main way of deploying the notion of relevance will be outlined in regard to the "group B" thinkers who, following an Aristotelian conception, describe relevance as a measure of personal significance.

2.1.1 Relevance in Information Retrieval

The study of information retrieval (IR) predates search engines, existing in various subdisciplines such as bibliometrics, Library and Information Science, Documentation Science, established in the twentieth-century, as well its more informal roots in the rise of scientific journals of the Seventeenth Century and wider concerns of epistemology and organisation. Stefano Mizzaro's exhaustive study "Relevance: The Whole History" represents a literature review of 160 papers, tracing these specifically twentieth-century developments as well as taking account of the wider history and looking to the future of IR. Mizzaro's survey and analysis outline the context into which Google entered, enhanced by the serendipitous date of publication, September 1997, which was the month that Brin and Page registered the domain google.com. Mizzaro's history outlines the centrality of the term relevance, while demonstrating that although the term has been adopted as the main metric for the discipline there is little agreement on what it represents. Mizzaro's opening epigraph borrows from an earlier literature review by Schamber et al. that comes to the same conclusion. Schamber et al. argue that,

> Since information science first began to coalesce into distinct discipline in the forties and early fifties, relevance has been identified as its fundamental and central concept […] an enormous body of information science literature is based on work that *uses* relevance, without thoroughly understanding what it *means.* (qtd. in Mizzaro 810, emphasis in original)

The term relevance has come to stand in for a multitude of different measurements. Mizzaro places great importance on a "widely recognized […] landmark in relevance

history" (815), that of Brian Campbell Vickery's presentation of two papers at the 1959 International Conference for Scientific Information (ICSI). In these two papers, Vickery established a "distinction between 'relevance to a subject' (the relevance of a document to a query for what concerns the topical component) and 'user relevance' (that refers to what the user needs)" (816), a distinction that maps onto the sectioning of this chapter between group A and group B thinkers, respectively. Vickery favoured the first metric, which became dominant in IR. By focusing on the relationship between subjects and queries, rather than subjects and users, mathematical models could be instituted with which to measure the effectiveness of retrieval systems. These choices, of the object of study and methods of measurement, enabled a wider discourse that facilitated a fundamental or objective measure of relations.

Søren Brier's *Cybersemiotics: Why Information is Not Enough!* details the influence of such an attitude, in reference to Vickery's later work *Information Science in Theory and Practice*, co-authored with Alina Vickery, which became one of the main textbooks for IR and its related disciplines. I quote Brier at length because he clearly, and critically, articulates the position of Vickery and Vickery as it stands in relation to the wider context of computer science in a manner that usefully describes the connected notions of the thinkers I have been describing as Group A:

> Vickery and Vickery define information science as the study of the communication of information in society. But their concept of information seems to be based on the statistical Shannon-Wiener concept of information and its mechanistic concept of a sender, a channel, and a receiver. This is combined with the cognitive science idea of information and information

> processing as a phenomenon unifying human, artificial, and natural systems. Vickery and Vickery contend that their concept of information encompasses all types of communication and causal connections and that information is as fundamental to reality as matter and energy […]
>
> Vickery and Vickery see nature as full of information. This is similar to the classical cybernetic perspective, in which information is understood as neg-entropy. According to this world view, natural objective information must have existed in the expanding universe before living beings and human minds. Information is more fundamental than either observers or interpreters. Accepting information as an objective, universal, law-determined thing that humans and machines absorb from nature, change, and multiply by thinking and by bringing it into society through languages, suggests that it must be possible to establish a unifying science of information. This development – along with the development of the computer, computer science, and the cognitive sciences – has […] promoted the idea of a unified information science for humans and living, physical, and artificial systems: the information processing paradigm (IPP). The major change here is that, as analysed and argued, there is no path leading from this rationalistic and physicalistic concept of information to a theory of signification and semantics. (418-419)

Such a conceptualisation of relevance, as more fundamental than subjective needs of users, has led to some useful measures of information systems. However, many – if not all – of these measurements are unfit for describing the web or search engines, as many of the metrics require an initial state, in which all the objects within the system are known. The vast majority of IR research that evaluates the relevance of

search engine results uses two measurements: precision and recall.[2] The metrics of precision and recall have continued to be used to evaluate search engines right up to the contemporary moment but are completely unsuited to the conditions of the web. The above studies tout numerical conclusions that hide the subjective criteria required when adapting existing statistical models to the context of the web. The continuation of such studies has reinforced the position that search engines can be judged according to an objective measure of relevance.

Rafiq Rather and S. M. Shafi's 2005 paper "Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology" is representative of the wider literature in the way that they define and deploy precision and recall. Their study is like many others and can be summarised in the following way. First, the authors decide on a set of queries on a particular topic, in this case biotechnology; the authors used topics taken from the subject headings of the Library of Congress Classification System, such as [cloning], [gene], ["monoclonal antibiotics"], and ["silage fermentation"], to use as queries. This method of choosing queries betrays the library origins of such a measurement and does not make any attempt to use queries that might reflect actual query formation, as demonstrated in the query log analysis studies of chapter one. The authors then establish a list of possible webpages they consider to be relevant for the search,

[2] Examples range from studies that predate Google up to those from the contemporary moment, see for example, in 1996 "Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology" (Chu and Rosenthal), "A Comparative Study of Web Search Service Performance" (Ding and Marchionini); in 1997 "Estimating the Recall Performance of Web Search Engines" (Clarke and Willett); in 1998 "On the Overlap, the Precision and Estimated Recall of Search Engines: A Case Study of the Query 'erdos'" (Bar-Ilan); in 2000 "The Evaluation of WWW Search Engines" (Oppenheim et al.); in 2005 "Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology" (Rather and Shafi); in 2009 "Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo" (Kumar and Prakash); in 2015 "The Precision and Recall of General Search Engines in Retreival of Images Related to Endocrine Diseases" (Hariri et al.); and in 2017 "Evaluating Search Effectiveness of Some Selected Search Engines" (Bute et al.), "An Algorithmic Query Refinement Model Based on Query Classification" (Ruban and Sam).

giving each a numerical value: four points for a research paper, three points for a patent, two points for an abstract of a conference paper, one point for a database entry for a book or paper, and zero points for anything else, including “company web pages, dictionaries, encyclopedia, organization, etc.” Their subdivisions for the scoring is fairly arbitrary and in general reflects their needs as researchers and their subjective expectations that search engines should work as glorified library catalogues; their scoring does not attempt to reflect the kinds of documents usually found on the web. Precision is then calculated by dividing the sum of the scores by ten, the number of results usually given on a search engine results page. The authors constrain their results to the first page only as they argue that most users do not consider results from page two, which on average is true, however, given the specific nature of their queries and the presumed context of the user being a biotechnology researcher looking for particular kinds of results, such a user may well travel beyond the first page of results. They then calculate recall, which is usually defined as equal to the total number of relevant documents retrieved, divided by the total number of relevant documents in a given collection. As the authors note, there “is no proper method of calculating absolute recall of search engines as it is impossible to know the total number of relevant [documents] in huge databases”. Therefore, they borrow from Clark and Willett 1997, who define a measure of relative recall to use with search engines by performing the query search on a number of different search engines, pooling all relevant results and then measuring how many of this total number each search engine’s first page provided. Not only does this exacerbate the issues with using only the first page of results – the specific results from the relative pool may well have been the eleventh result – but it ends up pitting a range of general purpose and specific search engines against one another. This

leads Shafi and Rather to compare AltaVista, Bioweb, Google, Hotbot, and Scirus with one another without acknowledging that Bioweb and Scirus – now discontinued science-specific search engines – are designed with the express purpose of giving results that are different from general search engines such as AltaVista, Google and Hotbot. This critique should not be considered as a specific attack on Shafi and Rather, instead, it seeks to demonstrate how traditional metrics for studying IR are unfit for search engines.

There are other metrics that researchers have used in attempts to avoid the pitfalls of relevance, such as Abbe Mowshowitz and Akira Kawaguchi's measurements of *bias*, through a predetermined "ideal distribution of items in a response set" (1194), however, this "ideal is approximated by the distribution produced by a collection of search engines" (1194). Therefore, even though the technique for measuring bias uses different mathematics to relevance, it still depends on producing a normative value, which only takes into account the highest ranked pages currently listed by search engines. As Mowshowitz and Kawaguchi describe, a "search engine is being weighed against its peers, not against an absolute norm derived from features of the universe. It might be desirable to adopt the latter approach, but it just is not feasible given the enormous size of the World Wide Web" (1194). Even though they accept that their measure "is a relative concept" (1194) the fallacy that bridging the gap from subjectivity to objectivity simply depends on technological improvements haunts their paper, as it does many others.

Relevance has become a key term when discussing search engine results. Technologist Adam Raff, in a 2009 *New York Times* op-ed, used the ongoing discussions regarding Net Neutrality, to highlight the increasing influence of search engines, he wrote that,

> Today, search engines like Google, Yahoo and Microsoft's new Bing have become the Internet's gatekeepers, and the crucial role they play in directing users to Web sites means they are now as essential a component of its infrastructure as the physical network itself. The F.C.C. [Federal Communications Commission] needs to look beyond network neutrality and include "search neutrality": the principle that search engines should have no editorial policies other than that their results be comprehensive, impartial and based solely on relevance.

Raff is certainly not using the term relevance in the sense of personalisation, which will be discussed shortly; instead, he uses the term as many others have, including Google themselves, to stand in for some sort of objective techno-deterministic measure that stands outside of subjective bias. Raff's statement reflects the importance of expectations, in line with the Heideggerian model outlined earlier, of keeping technologies ready-to-hand; in essence, Raff is asking for consistency and for search engines to surprise as little as possible.

The importance of expectations of consistency is further exacerbated by the metric of relative recall. Relative recall is a measure of consistency across all available search engines; such a measure is not negative in itself, it can be a useful tool for investigating antitrust allegations. This metric was most likely used in the EU's European Commission antitrust challenge, led by Margrethe Vestager, which on 27/06/2017 fined Google €2.4bn because its price comparison tool was deemed to be ranked unfairly. Vestager said that

> Google's strategy for its comparison shopping service wasn't just about attracting customers by making its product better than those of its rivals. Instead, Google abused its market dominance as a search engine by promoting its own comparison shopping service in its search results, and demoting those of competitors (qtd. in Chopra).

In addition, Vestager later commented that the "decision that Google is dominant gives us another starting point to look at other verticals, be it images, maps and local [advertising]" (qtd. in Harris). However, the difficulty is that a significant part of Google's success is due to the wealth of data and financial resources their market dominance affords them; Google's revenue alone should enable them to produce different, one might even hazard *better*, results than their competitors. The legal and economic complexities of the changing nature of monopolies in a digital context is too large to be dealt here in passing.[3] However, the example of the EU antitrust case does highlight the importance of relevance in a range of contexts, one of which being the comparisons between Google's results and its competitors.

Before moving on to the next section, addressing the Group B version of relevance, one further example needs to be highlighted, which problematises the usefulness of precision and relative recall for evaluating search engines even further. Not only have Google's competitors been shrinking in number, but those that are still active do not necessarily produce an alternative set of results. Outside of countries such as China that have their own search monopolies, the two alternatives, in terms of market share, are Bing and Yahoo!, with around 7% and 5% respectively.

---

[3] Siva Vaidhyanathan's *The Googlization of Everything: (And Why We Should Worry)* (2011) addresses the legal aspects of Google's role directly and Nick Srnicek's more recent *Platform Capitalism* (2017) provides a wider economic analysis of Google and other digital platforms such as Facebook, Airbnb and Uber. In addition, the final chapter of this thesis also addresses some of these issues in the context of advertising.

However, although it is not necessarily widely known, Yahoo!'s search engine is powered exclusively by Bing, the only difference between the two sites being that Yahoo! still acts as a portal for other curated sites, as described in chapter one. In addition to this, there is conclusive evidence to show that there have been examples where Bing has copied Google's results, rather than using their own metrics to provide search results. In 2011, Google ran a number of tests in which they fixed a range of nonsense queries, such as [delhipublicschool40 chdjob] and [hiybbprqag], to specially chosen unrelated webpage results. In doing so, they created a system in which copied results could be easily detected; such a system is similar to older methods to detect plagiarism used in cartography, such as the insertion of fake or paper towns, and fictitious entries into encyclopaedias. The tests were outlined in Sullivan's 2011 article "Google: Bing Is Cheating, Copying Our Search Results" and Amit Singhal's official Google blog post "Microsoft's Bing uses Google search results—and denies it". Putting aside the numerous legal issues that the incident raised, this example demonstrates how Google's results represent a gold standard so much so that its competitors simply want their results to replicate Google's. Such a situation means that traditional metrics for calculating relevance, in a non-personalised sense, cannot be applicable. Therefore, the rhetoric around neutral, objective or unbiased results is a rather strange fiction.

2.2 Group B: Search Is Undemocratic, Relevance Is a Measure of Personalisation, and All Answers Are Inherently Biased

Even before Google established a monopoly, there were many voices that dissented from the optimistic perspective that the web was inherently democratic. Douglas

Schuler's 1998 conference paper "Reports of the Close Relationship Between Democracy and the Internet May Have Been Exaggerated: Challenges and Opportunities for Rapprochement" develops the argument that in contrast to the views presented above, the web, even before the influence of search engines, was constrained by established institutions and a nexus of opaque power relations. The key point that Schuler argues is that the use of the word democracy is often used simplistically and in a way that misses what he sees as key aspects of democracy, in particular, that

> democracy requires a deliberative public process. This point contains three critical ideas: *deliberative* – adequate time must be allotted for hearing and considering multiple points of view; *public* – the discussion takes place in the daylight where it can be observed by all; and *process* – the procedures through which concerns are brought up, discussed and acted upon are clear and widely known. (emphasis in original transcript)

Although Schuler is not directly referring to search engines, deliberative democracy is imperative for our current context. If search engines have fundamental characteristics, I would argue that one of these is a hidden criterion of judgement: search engines function by making decisions on our behalf. This notion sets the stage for Group B's definition of search as personalised, biased and subjective. The importance of outlining these two different approaches lies in their incompatibility; both groups argue that their version of search represents the ideal nature of what search engines should aspire towards. If there can be any deliberative dimension to search engines, it must come in an open dialogue regarding the ideals of search

engine technology. Because the terminology, such as relevance, is shared it is crucial to understand the perspectives that particular technologists and commentators bring to these wider debates. After this section I will address the rather limited number of current academic studies that aim to evaluate search engine results, as well as outlining my own tests. The perspectives that academics take in such studies, their criteria and methods of evaluation is far more important than any particular set of results. What lies at the heart of search engine research, particularly in the social sciences or any research that aims to provide any kind of measure, is that there is no consensus on how search engines ought to function. This section, discussing Group B perspectives, aims not to collapse such disagreements, but to foreground them, in order that a more productive discourse can take place.

John Battelle's 2005 landmark study *The Search*, closes with a chapter devoted to what Battelle terms "perfect search" (251-280). In doing so, Battelle outlines what he considered to be the fundamental characteristics of search engines, compared with other ways we have organised knowledge throughout history, and what a fully-realised implementation of their logic would look like. Battelle describes "perfect search" in the following way:

> Imagine the ability to ask any question and get not just an accurate answer, but your perfect answer – an answer that suits the context and intent of your question, an answer that is informed by who you are and why you might be asking … While it's true that most questions don't have an objectively perfect answer, perfect search would provide *your* perfect answer, as you determine it – in a report form, perhaps, or by summarizing key points of view and trends. This perfect search also has perfect recall – it knows what you've

> seen, and can discern between a journey of discovery – where you want to find something new – and recovery – where you want to find something you've seen before […] the search engine of the future [is] more like an intelligent agent – or as Larry Page told me, a reference librarian with complete mastery of the entire corpus of human knowledge. (252)

The history of Autocomplete certainly showed a movement from general word prediction to contextual, situational suggestions that respond to previous searches of particular users. The examples of Google Now and 'the feed' also reproduce this kind of logic. Google's attitude to contextual advertising also follows a personalised model. Whether or not Google results are *personalised* is difficult to measure, although it will be discussed shortly. Battelle goes on to argue that such personalisation is "the stated goal of nearly every player in search, be it IBM, Microsoft, Google, Yahoo, or scores of others" (253). Battelle's study, in 2005, predates many of the dominant technology institutions that exemplify personalisation; in 2006, Twitter was launched and Facebook was opened up for public access, two events that fit into a wider trend towards algorithmically *relevant* content curation, the implications of which are returned to in chapter five.

This notion of subjective results has framed many studies that seek to evaluate search engine performance. Joshua Wright argues that fundamentally "bias […] is both necessary to and inherent within any useful indexing tool" (3). In "Personalized Search" Pitkow et al. also foreground the notion of personalisation as central to an alternative definition of relevance. Published in 2002, "Personalized Search" documents a series of search engine comparison tests, carried out two years prior, between existing search engines and the team's system of modifying

Google's results: Outride. Their Outride technology used various pieces of information, such as the "last 1,000 unique clicks of each user" (53), to re-rank Google's results for each user in a personalised manner. This, they argue in their paper, shifts the mode of relevance in a way that, using their metrics of evaluation, gave preferable outcomes. They describe the attitude embodied by Outride, in the following way:

> Focusing on the user enables a shift from what we call "consensus relevancy" where the computed relevancy for the entire population is presumed relevant for each user, toward personal relevancy where relevancy is computed based on each individual within the context of their interactions. The benefits of personalized search can be significant, appreciably decreasing the time it takes people – novices and experts alike – to find information. (50)

Their experiments tested participants in laboratory conditions by giving individuals an information retrieval task, asking them to use various search engines as well as a test condition in which Outride had re-ranked results, in order to be personalised around each user. Pitkow et al. measured the time it took for users to find the pieces of information. Given the Query Log Analysis of chapter one, their methodology should be addressed with some scepticism; retrieval of specific pieces of information has never been a representative activity of the usage of search engines. However, measuring search conditions in terms of seconds, allows them to draw numerical conclusions: on average users took 38.9 seconds when using Outride, whereas it took users 75.4 seconds (93.7% slower) with Google, 81 seconds (107.9% slower) with Yahoo!, 83.5 seconds (114.5% slower) with Excite, and 89.6 seconds (130.2%

slower) with AOL (54). Although we might question the value of these figures and suggest that a discussion of the impact of personalisation has wide-reaching consequences that cannot be captured with such metric, Google were impressed with the findings and consequently acquired the “intellectual property, including patent rights, source code, trademarks, and associated domain names, from Outride Inc.” (Google “Google Acquires Technology”) in 2001, a year before Pitkow et al. published their results. Although this acquisition means that many of the technical details of the study are not public information, it does indicate that Google were interested in incorporating this notion of personalised relevance in their search engine from at least 2001.

In 2005, a joint MIT and Microsoft search team, Teevan et al., followed up on the research by Pitkow et al., by establishing a model of personal relevance that is contrasted to relevance as established from the overall corpus. They focused on basic information retrieval tasks, which makes evaluation much easier, as the “users were asked to describe their intent and to rate the relevance of documents relative to their intent” (452) on a scale of “highly relevant, relevant, or not relevant to the query” (452). Some of the queries were self-selected and others were provided, such as “cancer”, “Bush”, “Web search”; the key aspect, however, is that relevance is not simply based around personalised relevance, but also linked to a specific intention. This becomes a problematic metric for any general purpose searches designed to be open-ended. Therefore, when Teevan et al. argue that “personalization algorithms can significantly improve on current Web search” (449), it is clear, in the context that this chapter has so far outlined, that *improve* represents an underlying attitude towards a particular conception of search. Teevan et al. are much more explicit

about their method of personalisation than Pitkow et al., outlining two models of personalisation:

> For a Web search engine to incorporate information about a user, a user profile must either be communicated to the server where the Web corpus resides or information about the results must be downloaded to the client machine where a user profile is stored. We have focused on the latter case, on re-ranking the top search results locally, for several reasons. For one, such a methodology ensures privacy; users may be uncomfortable with having personal information broadcast across the Internet to a search engine, or other uncertain destinations. Second, in the re-ranking paradigm, it is feasible to include computationally-intensive procedures because we only work on a relatively small set of documents at any time. Third, re-ranking methods facilitate straightforward evaluation. (450)

The three reasons as to why Teevan et al. follow local re-ranking are telling insights into how such decision-making occurs. First, they address potential user reaction to privacy issues; Teevan et al. do not consider the wider philosophical consequences of pragmatic reasoning, or the political implications of tailoring information flow, rather they address how users might feel “uncomfortable” (450). Second, re-ranking is less computationally-intensive, meaning that it would be cheaper to run and the results would be delivered more quickly. As has already been addressed, speed is a primary metric for Google. Finally, Teevan et al. prefer re-ranking because it allows for more straightforward evaluation. This again, speaks to the open-ended potential

of search results; having subjects give relevance scores for the same pool of results gives more measurable results than two completely different sets of results.

Using the linear scheme of evaluating search results on a scale from not relevant to highly relevant hides the nuanced differences between universal results and personalised ones. Law professor Eric Goldman reflects this approach in his 2006 article “Search Engine Bias and the Demise of Search Engine Utopianism” where he argues that “technological evolution will moot search engine bias” (198). Goldman conceives of bias in terms of results that cater to the most popular needs of users, and in turn, deliver “suboptimal results for searchers with minority interests” (198). If results are relevant to individuals then, Goldman argues, bias is not a problem because through data collection the results can do “a better job of searcher mindreading” (198). This reiterates the logic established by Pitkow et al. and Teevan et al., that relevance is a measure of matching results to the presumed intentions of specific users, the opposite conception to the notion of relevance outlined by the Group A thinkers. Goldman concludes by arguing a rather deterministic position that the results of search engine providers should not be regulated and that as technology improves so will user satisfaction, and therefore users would never see that their results are biased because they would always reflect their own sentiments:

> Complaints about search engine bias implicitly reflect some disappointed expectations. […] Fortunately, search engine bias may be largely temporal […] search engines naturally will continue to evolve their ranking algorithms and improve search result relevancy – a process that, organically, will cause the most problematic aspects of search engine bias to largely disappear. To

> avoid undercutting search engines' quest for relevance, this effort should proceed without regulatory distortion. (199-200)

As I have noted repeatedly in this chapter, the rhetoric around the *ideal* or *inevitable* logic of search engines is crucial as it sets the agenda for how search engines function. This is particularly important in the context of privacy and surveillance. Michael Zimmer's 2008 article "The Gaze of the Perfect Search Engine: Google as an Infrastructure of Dataveillance", published in the influential edited collection *Web Search* that he co-edited with Amanda Spink, discusses the extent to which search engines require personal data to function. Zimmer builds on Battelle's notion of *perfect search* to argue that relevant results require "the collection of personal information [as] a prerequisite of participation" (93). Dataveillance is, for Zimmer, a necessary part of the deal that users strike upon using Google's search engine, which he characterises in the following way:

> a Faustian bargain emerges with the quest for the perfect search engine: The perfect search engine promises breadth, depth, efficiency, and relevancy, but enables the widespread collection of personal and intellectual information in the name of its perfect recall […] One avenue for changing the terms of the Faustian bargain is to enact laws to regulate the capture and use of personal information by Web search engines. A recent gathering of leading legal scholars and industry lawyers [Zimmer is referring to the 2005 "regulating search: a symposium on search engines, law, and public policy" held at Yale Law School] to discuss the possibility of regulating search engines revealed,

> however, that viable and constitutional solutions are difficult to conceive, let alone agree upon. (93-94)

Zimmer's perspective is that in order for search engines to become "perfect" data collection and surveillance are necessary. While these topics are not a major focus of this thesis they are key questions for the study of digital culture more widely. Attitudes, such as Zimmer's, that view some search engine characteristics as unavoidable can end up downplaying significant social issues. Zimmer calls for greater transparency that would enable users to view the data Google stores on them. Google added such a feature in June 2016 called "My Activity", which shows users all the information stored about them. Although a useful step in the direction of transparency, such a tool does not challenge the more vital issue of *how* search engines should provide results. Whether search engines should personalise or aim for a general relevance is a much bigger issue than each user's privacy.

3.0 Academic Methodologies Shape Conclusions

Because there is such disagreement on the basic terms of debate, in particular the use of the word relevance to represent a number of contradictory viewpoints, research that aims to evaluate search results is often designed around the norms that each researcher brings, concerning how search engines *should* produce results. Given the constraints of search engine research, such as the lack of reproducibility, that Hargittai, as quoted at the beginning of this chapter, argues places search engine evaluation outside the basic tenets of science, the methods of evaluation have substantial impacts. Surveys, interviews and personal observation make up a large proportion of academic evaluation of search engine results: see for example, “Web Search Behavior of Internet Experts and Newbies” (Hölscher and Strube), “Beyond Logs and Surveys: In-Depth Measures of People's Web Use Skills” (Hargittai), “A Framework for Studying Differences in People’s Digital Media Uses” (Hargittai), “Internet Searchers Are Confident, Satisfied and Trusting – but They Are Also Unaware and Naïve” (Fallows), “Search Engine Use 2012” (Purcell et al.). Such studies prioritise user satisfaction and consequently, personalised relevance. It is rare for such surveys to ask if users thought the results were fair or objective, compared to whether their informational need was met. Fry et al. provide the only exception in “Search Engines and Expertise About Global Issues: Well-Defined Landscape or Undomesticated Wilderness?” (2008), the research of which consisted of interviewing academic experts in various fields to find out “whether the use of online resources enhances or diminishes the range of available sources of expertise” (255) to which they concluded that there are a number of different gatekeepers for different areas, but no “winner-takes-all” dominance across various domains. The

methodology of Fry et al. sets the study clearly within Group A described above, where results were judged on the basis of whether or not experts thought they reflected a comprehensive perspective on a topic.

The mathematics and computer science literature that considers search results is large; “An Overview of Search Engine Evaluation Strategies” (Goel and Yadav) provides a comprehensive literature review, although the papers considered do not challenge notions such as relevance or bias and instead formulate their research questions in terms of computational efficiency. This is not to generalise that computer scientists do not understand the difficulties of such metrics. For example, “Using Global Statistics to Rank Retrieval Systems Without Relevance Judgments” (Shi et al.) and “Automatic Performance Evaluation of Web Search Systems Using Rough Set Based Rank Aggregation” (Ali and Beg) both outline approaches for evaluating research results without relying on relevance as a metric. However, Dirk Lewandowski and Nadine Höchstötter’s survey the IR methods used in evaluating search results concludes with the statement that

> Today, nobody knows the real performance or accuracy of search engines. There are several studies dealing with a single aspect of quality measurement, but none that tries to evaluate search quality as a whole. There was a lack of an overview of empirical results and of quality measures to be used. (336)

The introduction to Lewandowski’s later edited collection of 2012 opens with a discussion concerning the lack of studies that take a general approach, stating that “it is clear that Web search engine research is still in its infancy” (“New Perspectives”

13). Therefore, the next section of this chapter seeks to build on some of the issues already foregrounded, by analysing a selection of results. As results vary due to a number of different reasons, this analysis will attempt to highlight those that seem to have the largest effect on search results. Google's results change throughout time; for this reason, a longitudinal study would highlight such changes. Using a single set of results, as is presented below, to speak to wider generalisations is difficult. However, I aim to provide a kind of close reading that opens up possible routes for analysis, rather than leading with a specific hypothesis. There are too few studies that systematically collect search results and analyse any patterns that emerge, rather than starting out with a specific agenda. Paul Reilly's 2008 article investigates the visibility of particular Northern Irish terrorist groups in search results and contains some broader questioning as to the general functions of search results. However, the paper focuses too heavily on comparing specific metrics, such as total returned results, that have little to do with the kinds of results actually produced. Susan Gerhart's "Do Web Search Engines Suppress Controversy?" provides a relatively broad analysis of search engine results through an evaluation of how easy or difficult it is to find specific controversies in search results; in listing the experimental limitations of such a study, Gerhart outlines the need for a method that discusses results without a specific research question. The section that follows attempts to provide a strategy that is separate from the methodologies of the social sciences in order to use search results to open-up discussion, rather than provide evidence for or against a specific claim.

4.0 Altering Variables to Change Search Results: An Experiment

This following section details a set of experiments carried out on 24 November 2015 in which a set of queries were searched a number of times, while adjusting certain variables in order to replicate different contexts (locations, languages, phrasing) to see how such differences would alter the search results. Such examples can shed light on the previous discussion of relevance and the tension between generalised and personalised results. The conclusion of the experiments was that queries are provided different search results when searching the same and related queries across regions, languages, and with different word choices for each search term. In addition, the results reveal that the variations in search results across different contexts are complex and hard to predict; in some contexts, particular characteristics – such as location – have a minimal effect, whilst in a different context – for example, when searching in another language – that same change of location might have a much more significant impact on the results. Google Search takes into account a large number of different signals every time a query is submitted. Although the experiment below only addresses three of these signals, even this small study shows the level of variation and consequently how much is at stake when results are tailored for individual users.

Google have clear legal responsibilities to the laws of specific countries. Their Transparency Report details the instances where individuals, companies or countries have demanded they remove certain results and the decision taken in response. For example, “11% of the German removal requests are related to pro-Nazi content or content advocating denial of the Holocaust, both of which are illegal under German

law" (Google "Transparency Report, Germany"). However, unless they are legally compelled to comply, Google claim not to censor results; for example, another German appeal details "a request from a law enforcement agency to remove a YouTube video that allegedly depicts state symbols in a disparaging way" to which Google "did not remove the video". Localised search reflects local laws, but aside from these explicit cases of censorship, Google do not outline how different contexts change results, beyond their usual rhetoric of making results *relevant*. But search results *do* change, every time a query is searched Google base the results and their rank on specific signals, over different 200 factors, as outlined in chapter one.

The example used in this chapter, to foreground differences in search results, is a set of queries related to homosexuality; these were repeatedly searched under different contextual conditions, in order to highlight which signals had an impact on the results provided by Google. Three main criteria were tested: the geographical location that the queries were searched from, the language that the queries were written in, and the kinds of phrasing or tone used in the query. Homosexuality is a useful example for two main reasons. First, same-sex relationships are treated very differently in various countries, fully-accepted in some locations while representing a controversial issue in others. Second, the topic of same-sex relationships can be addressed through a number of different discourses that appeal to different kinds of evidence and argument. A significant proportion of webpages returned by Google for queries related to homosexuality aim to be persuasive and are written with an appeal to different kinds of authority, including legal, moral, religious, medical, and anecdotal perspectives. When analysing the differences between search results, it is not only important to highlight whether webpages present a positive or negative attitude toward homosexuality but also the form that these arguments take, or what

could be referred to as the *genre* of particular pages. When comparing the effects of word choice, for example comparing the different kinds of results for a query using the word [gay] to one that used the word [homosexual], what *kinds* of pages are represented? Would the Google results in some countries provide academic research papers, while others, listings for gay clubs? Would searching in one language provide personal accounts written by individuals, while another results in a high number of institutional pages deploying religious or legal arguments? The following study provides evidence that searching the same queries in different contexts significantly alters the kinds of search results returned, both in terms of their attitude towards homosexuality and the kinds of discourses deployed. In addition, the results of this case study demonstrate that Google's results were far from apolitical and have shifting outlooks, dependent on the situation of the user searching a query.

The study involved searching a small number of query variants related to homosexuality in English, French, and Arabic in the countries of the United Kingdom, France and Morocco.[4] These queries were split between general, single word open-ended queries, such as [homosexuality], and more specific queries phrased as questions, such as [is being homosexual wrong?]. Word choice was tested by searching each query using the word [homosexual] or the word [gay]. Phrasing was tested by comparing two ways of formulating a question, relying on either [wrong] or [good].[5] These queries were taken from Google's Autocomplete suggestions. Each query was searched in all three countries and in all three languages and the all the webpage results on the first page of Google results were followed and given a rating.

[4] Although all the searches were carried out in the UK, a Virtual Private Network (VPN) was used to send the queries from the required geographical search locations. In addition, each country-specific domain, google.co.uk, google.fr and google.co.ma was used when searching from that country.
[5] The query [is being homosexual right?] was originally chosen for comparison with [is being homosexual wrong], however, the wording of that query led to pages regarding gay rights and so did not work as a comparative antonym. Therefore, an alternative Autocomplete suggestion, [is being homosexual good?], was used instead.

The results were rated using two combined scales: first, whether the page was positive, negative or neutral towards homosexuality (listed as Pro, Con or Neutral – see fig. 5 for the complete list of queries and contexts). Second, the mode of discourse, or the *genre* of the page, was rated as either “Factual” or Opinion based. It should be made clear that the difference between the rating of some pages as Opinion and others as “Factual” was not in any way related to whether or not the pages were convincing, robust or credible. Although there were variations within these categories, there was a clear separation between pages that were framed as Opinion-based articles or blogs and others that framed their site as an encyclopaedia entry or medical diagnosis (see figs. 6-11 for an example of each category).

Country	Language	Search Query	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
UK	English	Homosexuality	1	4	0	5	1	3
France	English	Homosexuality	2	2	0	4	0	2
Morocco	English	Homosexuality	2	1	0	3	0	7
UK	English	Gay	3	7	0	1	0	0
France	English	Gay	6	2	0	1	0	0
Morocco	English	Gay	2	3	0	3	0	0
UK	French	homosexualité	6	1	0	2	0	0
France	French	homosexualité	6	1	0	2	0	0
Morocco	French	homosexualité	6	1	0	2	0	0
UK	French	gai	0	0	0	0	0	0
France	French	gai	1	0	0	6	0	0
Morocco	French	gai	1	0	0	6	0	0
UK	Arabic	الشذوذ الجنسي	0	0	0	3	3	4
France	Arabic	الشذوذ الجنسي	0	0	0	4	2	4
Morocco	Arabic	الشذوذ الجنسي	0	0	0	4	1	4
UK	Arabic	مثلي الجنس	2	3	0	4	1	0
France	Arabic	مثلي الجنس	2	4	0	2	1	0
Morocco	Arabic	مثلي الجنس	3	3	0	2	1	1
UK	English	Is being homosexual wrong?	0	1	3	1	0	5
France	English	Is being homosexual wrong?	0	2	1	0	0	7
Morocco	English	Is being homosexual wrong?	0	1	2	0	0	7
UK	English	Is being gay wrong?	0	3	2	0	0	5
France	English	Is being gay wrong?	0	4	2	0	0	4
Morocco	English	Is being gay wrong?	0	3	2	0	0	5
UK	French	Est-ce qu'être homosexuel mal?	5	4	0	1	0	0
France	French	Est-ce qu'être homosexuel mal?	3	4	2	1	0	0
Morocco	French	Est-ce qu'être homosexuel mal?	3	4	1	2	0	0
UK	French	Est-ce qu'être mauvais gay?	4	5	1	0	0	0
France	French	Est-ce qu'être mauvais gay?	2	5	2	1	0	0
Morocco	French	Est-ce qu'être mauvais gay?	3	4	2	1	0	0
UK	Arabic	يجترئ خاطئ مثلي الجنس	0	1	2	2	0	5
France	Arabic	يجترئ خاطئ مثلي الجنس	0	1	1	4	1	3
Morocco	Arabic	يجترئ خاطئ مثلي الجنس	0	1	0	4	1	4
UK	Arabic	يجري خاطئ مثلي الجنس	0	1	1	3	0	5
France	Arabic	يجري خاطئ مثلي الجنس	0	1	1	4	1	3
Morocco	Arabic	يجري خاطئ مثلي الجنس	0	1	1	3	2	3
UK	English	Is being homosexual good?	1	2	6	1	0	0
France	English	Is being homosexual good?	0	1	6	2	0	1
Morocco	English	Is being homosexual good?	2	3	4	0	0	1
UK	English	Is being gay good?	1	8	1	0	0	0
France	English	Is being gay good?	1	7	1	0	0	1
Morocco	English	Is being gay good?	1	6	2	0	0	1
UK	French	Est-ce qu'être bonne homosexuelle?	2	0	0	2	0	0
France	French	Est-ce qu'être bonne homosexuelle?	3	2	0	4	0	1
Morocco	French	Est-ce qu'être bonne homosexuelle?	3	2	0	4	0	1
UK	French	Est-ce qu'être bonne gay?	3	6	1	0	0	0
France	French	Est-ce qu'être bonne gay?	4	3	2	1	0	0
Morocco	French	Est-ce qu'être bonne gay?	5	1	3	1	0	0
UK	Arabic	يجري جيدة مثلي الجنس؟	0	2	0	1	0	0
France	Arabic	يجري جيدة مثلي الجنس؟	0	3	0	3	0	0
Morocco	Arabic	يجري جيدة مثلي الجنس؟	0	0	0	0	0	0
UK	Arabic	يجري مثلي الجنس جيدة؟	3	3	1	1	1	0
France	Arabic	يجري مثلي الجنس جيدة؟	2	0	0	1	0	0
Morocco	Arabic	يجري مثلي الجنس جيدة؟	1	3	2	2	0	1

Figure 5. This table shows the complete list of the 54 search queries tested and the ratings of the first page of results for each.

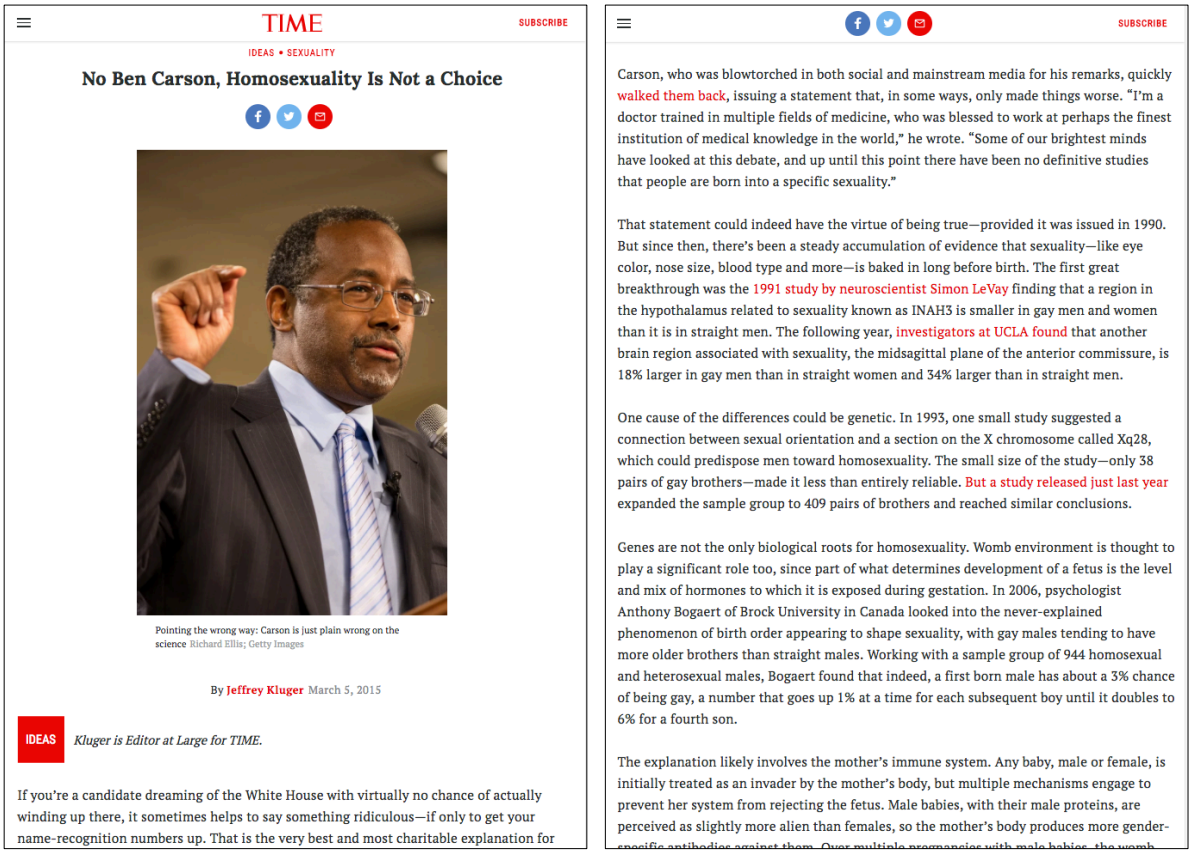


Figure 6. An example of a webpage result categorised as "Factual" Pro: "No Ben Carson, Homosexuality Is Not a Choice" Written by *Time*'s Jeffrey Kluger the article takes a clear pro-homosexuality stance. The page bases its arguments on reference and citation to hyperlinked scientific studies and is framed in a "Factual" discourse throughout (see Kluger).[6] Accessed 22/08/2017.

[6] These examples were added after the experiment was repeated in 2017 and therefore some of the pages were published after the original 2015 test. Figures 6-11 represent examples of the kinds of ratings given to pages, rather than exact results found in the 2015 study.

Figure 7. An example of a webpage result categorised as Opinion Pro: "Why Being Gay Is Better Than Being Straight" Written by Matt Stopera and Lauren Yapalater, two writers for Buzzfeed, the page is a humorous list of pop-culture references that takes a pro-homosexuality stance and does not cite any "Factual" evidence (see Stopera and Yapalater). Accessed 22/08/2017.

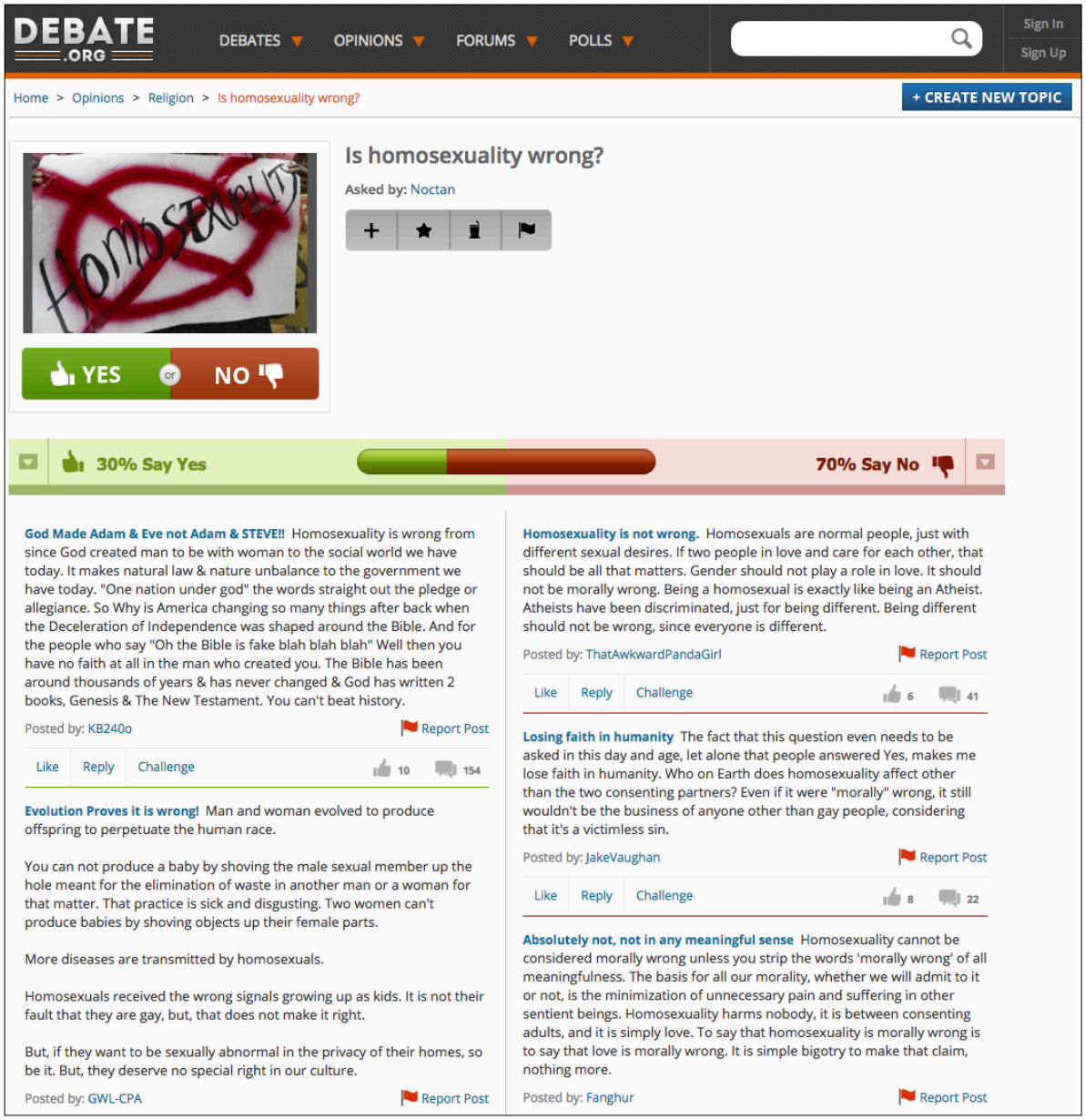


Figure 8. An example of a webpage result categorised as Opinion Neutral or Mixed: debate.org “Is homosexuality wrong”. This page is an open forum for users to submit their opinions on a particular topic. The page presents a mix of Pro and Con stances on homosexuality and the submitted posts do not generally cite any “Factual” evidence, rather they provide various personal opinions (see debate.org). Accessed 22/08/2017.

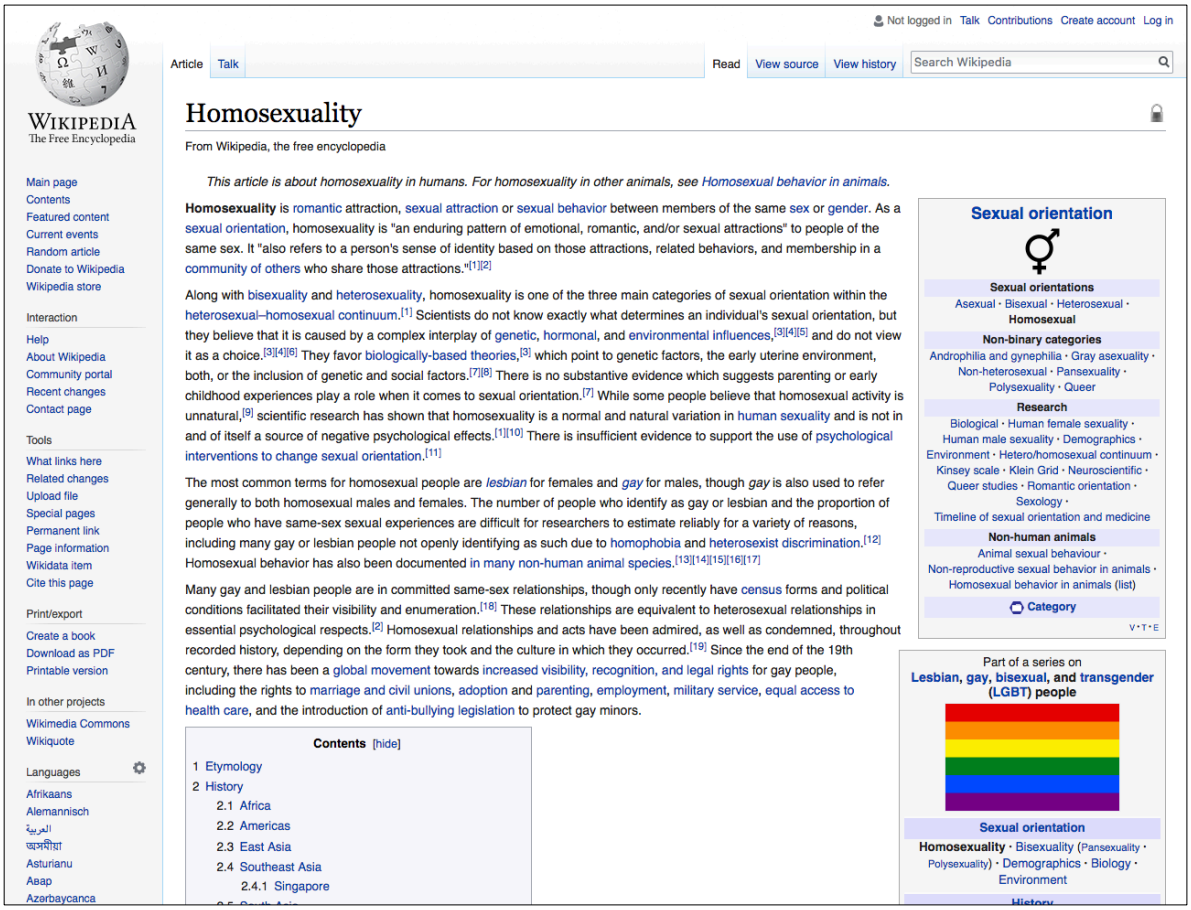


Figure 9. An example of a webpage result categorised as “Factual” Neutral: Wikipedia’s English entry for “Homosexuality”. One of Wikipedia’s key requirements for pages is that they are “verifiable against a published reliable source, thereby excluding editors' opinions and beliefs” (Wikipedia “About”). Reading through this entry confirms that it frames is discourse as “Factual” and does not take a clear Pro or Con stance regarding homosexuality (see Wikipedia “Homosexuality”). Accessed 22/08/2017.

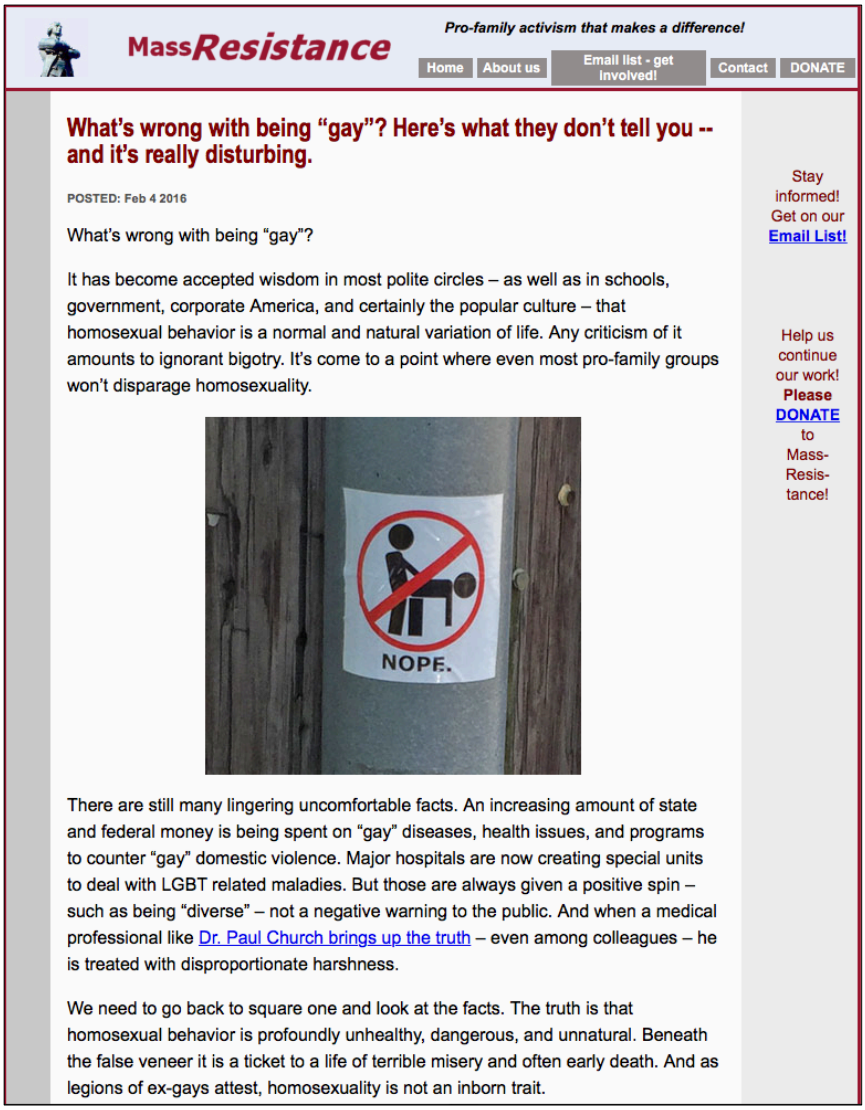
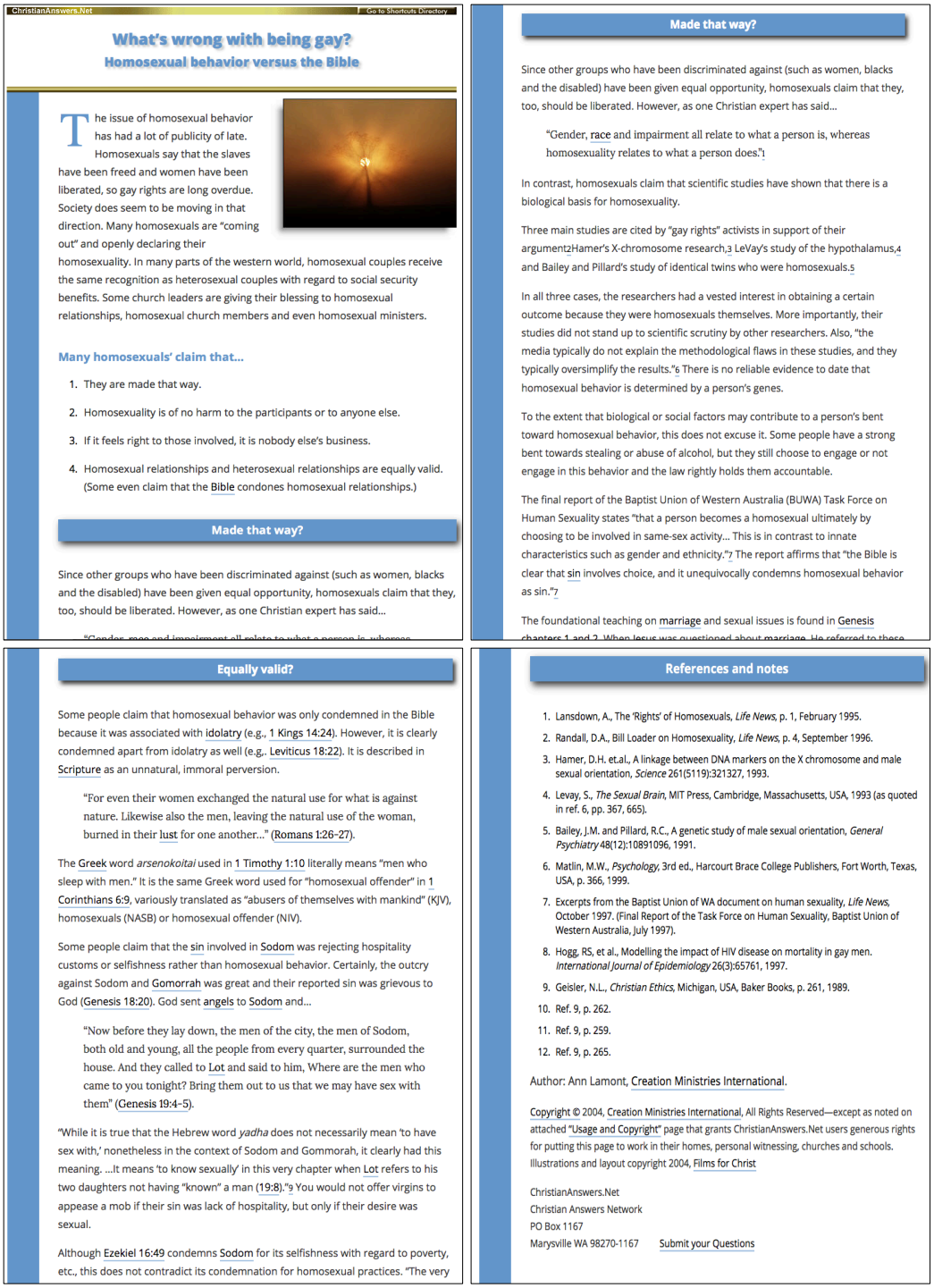


Figure 10. An example of a webpage result categorised as Opinion Con: Mass Resistance "What's wrong with being "gay"? Here's what they don't tell you – and it's really disturbing". This anonymously written page takes a clear anti-homosexuality stance and, although the article does use terms such as "truth" and "fact" throughout, it does not present references or citations to external sources as the basis for the claims made. The hyperlink "Dr. Paul Church brings up the truth", pictured above, appears as if it might indicate a "Factual" discourse, however, the link is to another article from the same website that details the author's personal opinion regarding a doctor who was fired for violating the discrimination and harassment policy of their workplace (see Mass Resistance). Accessed 22/08/2017.



ChristianAnswers.Net Go to Shortcuts Directory

What's wrong with being gay?

Homosexual behavior versus the Bible

The issue of homosexual behavior has had a lot of publicity of late. Homosexuals say that the slaves have been freed and women have been liberated, so gay rights are long overdue. Society does seem to be moving in that direction. Many homosexuals are "coming out" and openly declaring their homosexuality. In many parts of the western world, homosexual couples receive the same recognition as heterosexual couples with regard to social security benefits. Some church leaders are giving their blessing to homosexual relationships, homosexual church members and even homosexual ministers.

Many homosexuals' claim that...

1. They are made that way.
2. Homosexuality is of no harm to the participants or to anyone else.
3. If it feels right to those involved, it is nobody else's business.
4. Homosexual relationships and heterosexual relationships are equally valid. (Some even claim that the Bible condones homosexual relationships.)

Made that way?

Since other groups who have been discriminated against (such as women, blacks and the disabled) have been given equal opportunity, homosexuals claim that they, too, should be liberated. However, as one Christian expert has said...

Made that way?

Since other groups who have been discriminated against (such as women, blacks and the disabled) have been given equal opportunity, homosexuals claim that they, too, should be liberated. However, as one Christian expert has said...

> "Gender, race and impairment all relate to what a person is, whereas homosexuality relates to what a person does."[1]

In contrast, homosexuals claim that scientific studies have shown that there is a biological basis for homosexuality.

Three main studies are cited by "gay rights" activists in support of their argument[2]Hamer's X-chromosome research,[3] LeVay's study of the hypothalamus,[4] and Bailey and Pillard's study of identical twins who were homosexuals.[5]

In all three cases, the researchers had a vested interest in obtaining a certain outcome because they were homosexuals themselves. More importantly, their studies did not stand up to scientific scrutiny by other researchers. Also, "the media typically do not explain the methodological flaws in these studies, and they typically oversimplify the results."[6] There is no reliable evidence to date that homosexual behavior is determined by a person's genes.

To the extent that biological or social factors may contribute to a person's bent toward homosexual behavior, this does not excuse it. Some people have a strong bent towards stealing or abuse of alcohol, but they still choose to engage or not engage in this behavior and the law rightly holds them accountable.

The final report of the Baptist Union of Western Australia (BUWA) Task Force on Human Sexuality states "that a person becomes a homosexual ultimately by choosing to be involved in same-sex activity... This is in contrast to innate characteristics such as gender and ethnicity."[7] The report affirms that "the Bible is clear that sin involves choice, and it unequivocally condemns homosexual behavior as sin."[7]

The foundational teaching on marriage and sexual issues is found in Genesis

Equally valid?

Some people claim that homosexual behavior was only condemned in the Bible because it was associated with idolatry (e.g., 1 Kings 14:24). However, it is clearly condemned apart from idolatry as well (e.g., Leviticus 18:22). It is described in Scripture as an unnatural, immoral perversion.

> "For even their women exchanged the natural use for what is against nature. Likewise also the men, leaving the natural use of the woman, burned in their lust for one another..." (Romans 1:26-27).

The Greek word *arsenokoitai* used in 1 Timothy 1:10 literally means "men who sleep with men." It is the same Greek word used for "homosexual offender" in 1 Corinthians 6:9, variously translated as "abusers of themselves with mankind" (KJV), homosexuals (NASB) or homosexual offender (NIV).

Some people claim that the sin involved in Sodom was rejecting hospitality customs or selfishness rather than homosexual behavior. Certainly, the outcry against Sodom and Gomorrah was great and their reported sin was grievous to God (Genesis 18:20). God sent angels to Sodom and...

> "Now before they lay down, the men of the city, the men of Sodom, both old and young, all the people from every quarter, surrounded the house. And they called to Lot and said to him, Where are the men who came to you tonight? Bring them out to us that we may have sex with them" (Genesis 19:4-5).

"While it is true that the Hebrew word *yadha* does not necessarily mean 'to have sex with,' nonetheless in the context of Sodom and Gommorah, it clearly had this meaning. ...It means 'to know sexually' in this very chapter when Lot refers to his two daughters not having "known" a man (19:8)."[9] You would not offer virgins to appease a mob if their sin was lack of hospitality, but only if their desire was sexual.

Although Ezekiel 16:49 condemns Sodom for its selfishness with regard to poverty, etc., this does not contradict its condemnation for homosexual practices. "The very

References and notes

1. Lansdown, A., The 'Rights' of Homosexuals, *Life News*, p. 1, February 1995.
2. Randall, D.A., Bill Loader on Homosexuality, *Life News*, p. 4, September 1996.
3. Hamer, D.H. et.al., A linkage between DNA markers on the X chromosome and male sexual orientation, *Science* 261(5119):321327, 1993.
4. Levay, S., *The Sexual Brain*, MIT Press, Cambridge, Massachusetts, USA, 1993 (as quoted in ref. 6, pp. 367, 665).
5. Bailey, J.M. and Pillard, R.C., A genetic study of male sexual orientation, *General Psychiatry* 48(12):10891096, 1991.
6. Matlin, M.W., *Psychology*, 3rd ed., Harcourt Brace College Publishers, Fort Worth, Texas, USA, p. 366, 1999.
7. Excerpts from the Baptist Union of WA document on human sexuality, *Life News*, October 1997. (Final Report of the Task Force on Human Sexuality, Baptist Union of Western Australia, July 1997).
8. Hogg, RS, et al., Modelling the impact of HIV disease on mortality in gay men. *International Journal of Epidemiology* 26(3):65761, 1997.
9. Geisler, N.L., *Christian Ethics*, Michigan, USA, Baker Books, p. 261, 1989.
10. Ref. 9, p. 262.
11. Ref. 9, p. 259.
12. Ref. 9, p. 265.

Author: Ann Lamont, Creation Ministries International.

Copyright © 2004, Creation Ministries International, All Rights Reserved—except as noted on attached "Usage and Copyright" page that grants ChristianAnswers.Net users generous rights for putting this page to work in their homes, personal witnessing, churches and schools. Illustrations and layout copyright 2004, Films for Christ

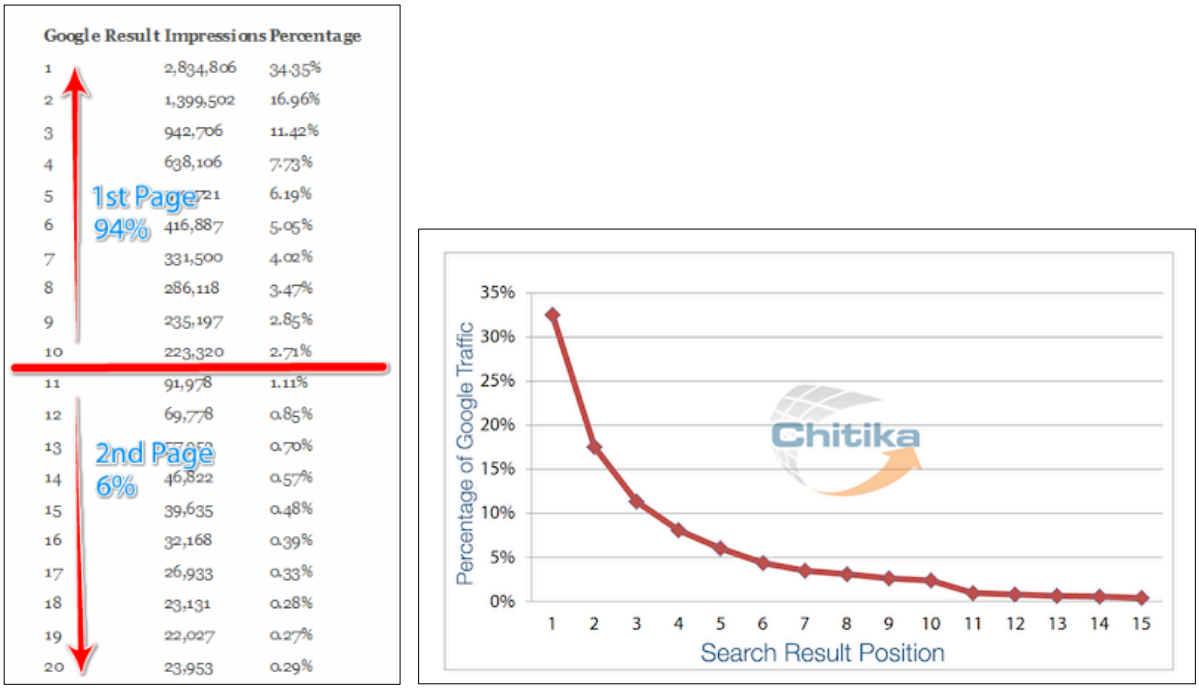
ChristianAnswers.Net
Christian Answers Network
PO Box 1167
Marysville WA 98270-1167 Submit your Questions

Figure 11. An example of a webpage result categorised as "Factual" Con: Christian Answer's "What's wrong with being gay? Homosexual behavior versus the Bible". The result takes a clearly Negative stance regarding homosexuality. The discourse of the page is framed as "Factual" as each statement made throughout is linked to a piece of external evidence, including biblical scripture and scientific papers (see Lamont). Accessed 22/08/2017.

All searches were carried out using a Virtual Private Network (VPN), which reroutes searches through a separate Internet Protocol (IP) address sent from a server in a particular country. In addition to enabling the queries to be sent from chosen locations (the UK, France and Morocco), using a VPN kept the searches anonymous, so that no personalised data, such as search history would be used. VPN's are usually used for three reasons: keeping web activity anonymous, accessing country-specific versions of sites (for example, gaining access to Netflix's US library of content while outside the US), or circumventing country-specific censorship. Because the Moroccan government has been known to censor Internet content, individuals using a VPN are more likely to be doing so to *avoid* searching from Morocco, rather than explicitly choose to search from there. Although there are many VPN services available, most do not have servers based in Morocco, Hide My Ass! is one of the only VPN services with servers in Morocco and was therefore used for searching all queries.

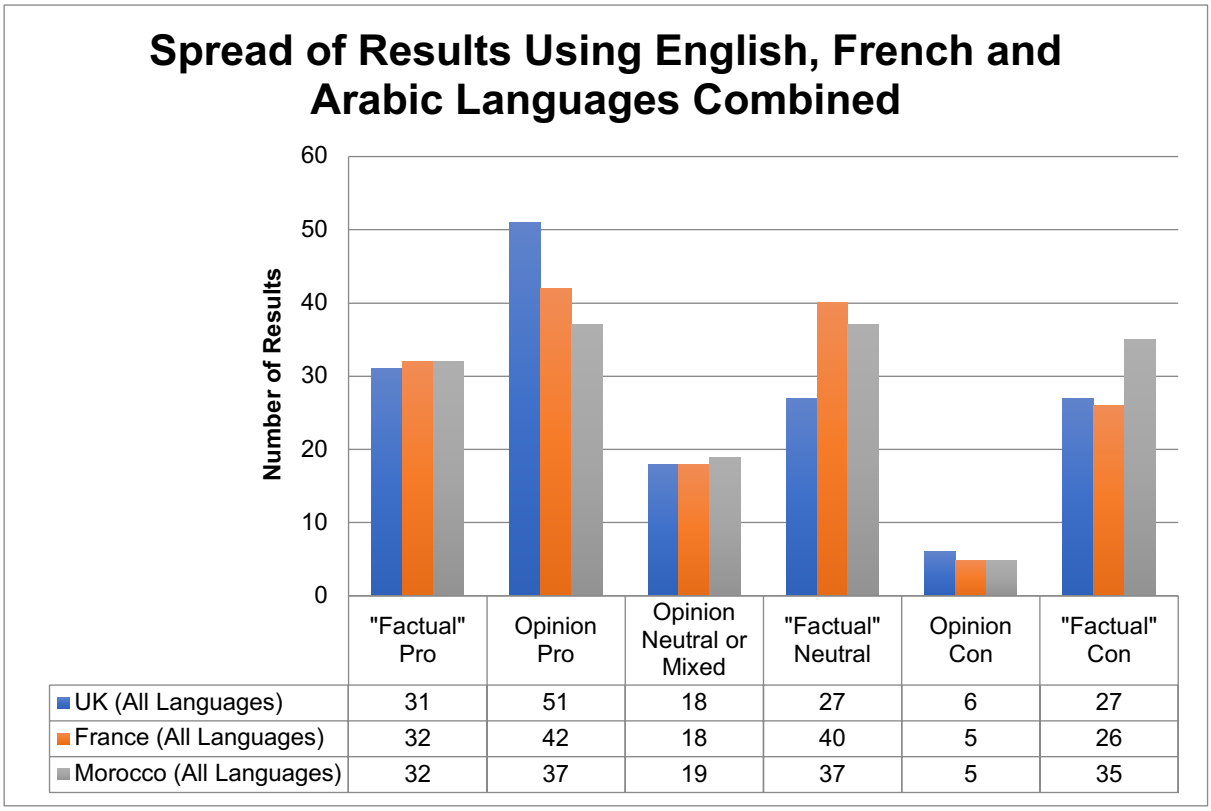
Only results from Google's first page were taken into account. As highlighted in chapter one, research by the advertising agency Chitika in 2013 (see figs. 12 and 13) shows that the number of results that users select on the second page of Google results is insignificantly low. Therefore, to draw on any results from the second page of any Google search results would be unrepresentative of a general search experience. The reason for drawing on all results from the first page, rather than the first three, four or five results, is to describe the overall landscape of search for a user. The first couple of results get a much larger proportion of users selecting them. However, I wanted the experimental design to reflect the range of positions provided in the first page, rather than just judging the top link. For this reason, all the links provided on the first page of results were deemed important. The user might select

the top result every time, but there is a significant difference between a results page in which all the results are ideologically in line with one another and a results page that represents a mix of viewpoints and discursive modes. Such a difference may have a significant impact on how the attitudes of an individual are challenged or reinforced.



Figures 12 and 13. Showing that users rarely navigate to Google's second page of results. Taken from Chitika Insights "The Value of Google Result Positioning"

The combination of the searches led to 54 separate queries, their distribution is shown below (see fig. 14).



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
UK (All Languages)	31	51	18	27	6	27
France (All Languages)	32	42	18	40	5	26
Morocco (All Languages)	32	37	19	37	5	35

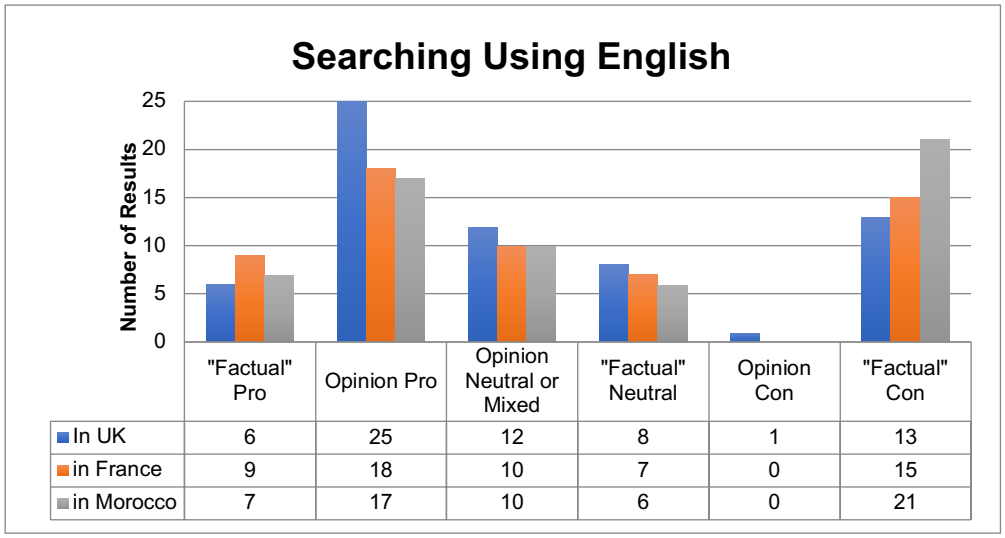
Figure 14. This graph shows the data from all three languages combined for each specific location.

The above graph shows a representation of the search results for all three languages searched in a particular country. The graph shows that there were some overall trends in the kinds of pages provided by Google, but also that there were differences between the kinds of results when searched in each country. For example, Positive Opinion-based pages occurred more often in the UK than in Morocco, whereas Negative “Factually” based arguments occurred more often when searching in Morocco. There were common trends shown across all countries when taking all three languages into account. For example, there was a greater number of Negative accounts of homosexuality framed in “Factual” modes of discourse, than Opinion-based arguments. Pages with Positive attitudes towards homosexuality, on the other hand, were more commonly Opinion-based in their discourse, but those

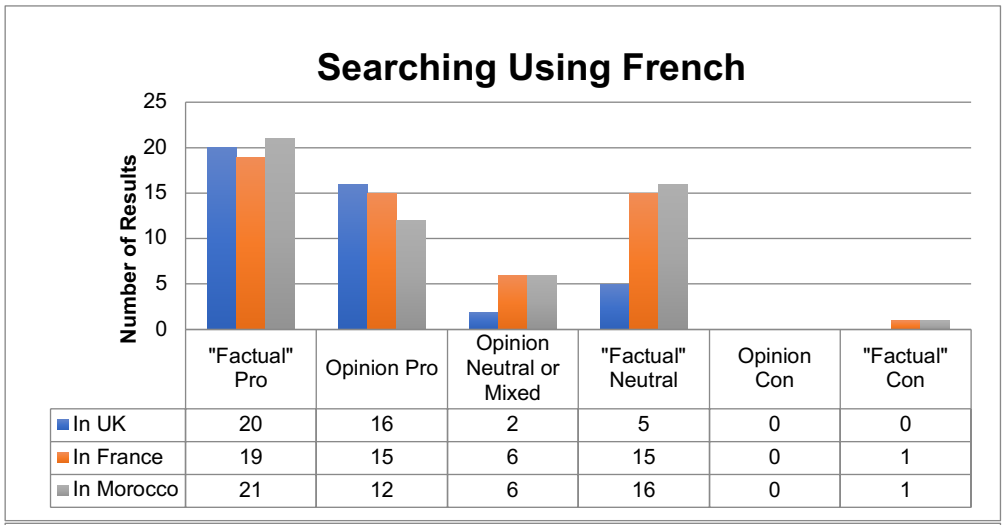
pages did exhibit a greater diversity of modes or genres than the Negative results. The observation that Positive representations occurred in a range of modes whereas Negative accounts appeared almost solely as cut-and-dried “Factual” statements, whether religious or scientific, that rule homosexual practices immoral or as an illness, is worth noting. Even if there had been a balance of Opinion and “Factual” in the Positive results, only getting “Factual” Negative results frames the issue of same-sex relationships within a wholly different discourse.

4.1 Comparisons of Google's Geographically Regional Domains

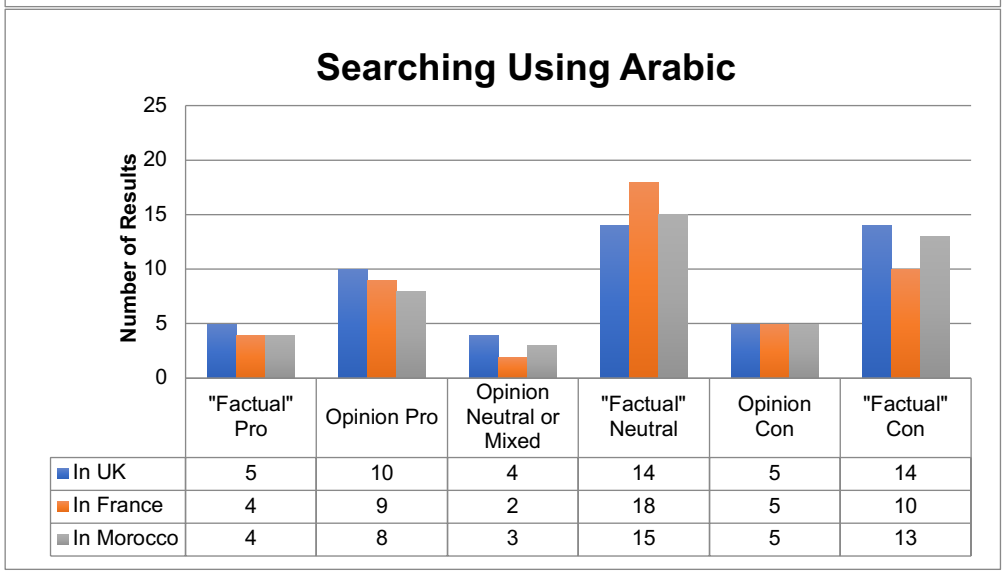
Figs. 15-17 show a graphical representation of the data split across language differences. The main conclusion to be drawn from these results is that language had a much bigger effect than geographical domain. As each cluster of bars represents different countries, any differences between them demonstrate the impact of geographic domain changes. It is clear that geography did not have a significant effect across each language; however, its impact varied for each context. Arabic, for example, had much less deviation between countries: searching in Arabic returned the same kind of results in each of the three countries tested. By contrast, search using English in different locations led to a greater level of variation in results. Searching using English in the UK provided a high number of Positive Opinion-based results, in comparison to searching using English in France, or English in Morocco. Searching using English in Morocco also produced more results with "Factually" framed Negative attitudes (Con) of homosexuality than using English in France or the UK. At the other end of the scale, searching using English in France produced more results that were Positive about homosexuality and framed in a "Factual" discourse. Searching using French was relatively consistent across domains except for results in the UK; searching using French in the UK led to very few results that were either Neutral/Mixed Opinions or "Factually" framed Neutral pages. Considering that when searching using French there were almost no Negative results – searching using French in France and Morocco led to both Positive and Mixed/Neutral results, searching using French in the UK led to an almost exclusively Positive set of results.



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
In UK	6	25	12	8	1	13
in France	9	18	10	7	0	15
in Morocco	7	17	10	6	0	21



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
In UK	20	16	2	5	0	0
In France	19	15	6	15	0	1
In Morocco	21	12	6	16	0	1



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
In UK	5	10	4	14	5	14
In France	4	9	2	18	5	10
In Morocco	4	8	3	15	5	13

Figures 15-17. These graphs shows the results for each individual language in across the three geographical domains.

A number of domain-specific conclusions can be drawn from such results. For instance, searches that used English were strongly affected by changes of location, whereas the search results of queries searched using Arabic were less affected by changes in location. The reasons for this might be based on the frequency of second language speakers. Perhaps the variation is due to a particular user base of individuals actually using English regularly in these various countries, which provides Google with a larger amount of data to draw from in order to make ranking decisions for these specific contexts. As Arabic is a less common second language than English, there are fewer users using Arabic to search in France and the UK. Other evidence for this claim lies in the different results provided for queries searched using Arabic in France and the UK and queries using French in the UK and Morocco (see figs. 16 and 17). For example, the percentage of Arabic speakers living in France is significantly higher than in the UK. As Emmanuelle Talon stated in 2012 in *Le Monde*, “With four million speakers, Arabic is the second most spoken language on French territory”; this amounts to roughly 6% of the population of France. In the UK, according to the 2011 census, there were 159,000 Arabic speakers; this roughly equates to 0.3% of the population. Considering the history of the French Protectorate in Morocco, immigration from the Maghreb region, and other related cultural associations, I predicted that searching using French in Morocco or using Arabic in France might demonstrate a greater overlap in the kinds of results produced, compared to searching using English in Morocco or Arabic in the UK. However, rather than similar results, searching using Arabic in France produced a bigger deviation from using Arabic in Morocco results than searching using Arabic in the UK (see figs. 17 and 16). Comparing the results from using Arabic in France with using Arabic in Morocco (see fig. 17) shows that there were three fewer “Factual” Con

results and three more “Factual” Neutral results in France than Morocco. The discourse, “Factual”, stayed the same when searching using Arabic in France or Morocco, but the attitude of the results presented was more Negative in Morocco and Neutral in France. This conclusion may play into particular prejudgements regarding how homosexuality is treated in France compared to Morocco; however, addressing other examples shows that each context – or combination of language, geography, word choice, and phrasing – does not lead to predictable outcomes.

The different criteria have complex and unpredictable relationships that do not work on an additive basis. An example that demonstrates this can be seen by comparing the differences in the number of Opinion Neutral or Mixed and “Factual” Neutral results of the searches using English in the UK and using French in the UK (fig. 15 and 16). Fig. 15 shows that the searches using English returned the largest number of Opinion Neutral or Mixed and “Factual” Neutral (i.e. results that explicitly did not take a distinct stance) when searched in the UK, compared with France and Morocco. If these criteria had an additive relationship, we would predict that using another language in the UK might follow a similar pattern. However, looking at the results of Opinion Neutral or Mixed and “Factual” Neutral when searching using French in the UK (fig. 16), the amount returned was significantly lower than in France or Morocco. Therefore, these different contexts do not have additive relationships, as if the French search results has been anglicised or UK-ified. There were a number of times when geography played a minimal role in the results returned, which is why results such as this one are noteworthy. Rather than seeing each characteristic as a form of filter, or as an ingredient in a recipe which produces results, each set of results is created through an assemblage of intersecting criteria. There are differences in the results for different countries and different languages, but these

are not predictable and do not follow a logical pattern whereby changing one specific metric has the same kind of result in all other contexts. Therefore, it would be wrong to make an overarching conclusion concerning a particular language, for example, that searching using Arabic promotes particular kinds of pages, because searching using Arabic has a very different impact in the UK than it does in France. However, comparing specific contexts demonstrates that Google results may fulfil very different roles in various contexts. Looking again at fig. 16, searching using French, shows, in aggregate, a very particular attitude presented. The lack of both forms of Con results highlights the kind of editorial role that Google might fulfil to French speakers. According to these results, French speakers searching in the UK would receive almost exclusively Positive results, whether framed "Factually" or as Opinion. Searching using French in the UK produced results that were undeniably pro-homosexual, whereas the results from searching using English in the UK *and* English in France (fig. 15) were much more Mixed and showed a higher number of "Factual" Negative results and few "Factual" Positive ones.

Based from these results, searching using English in each of the three countries (fig. 15) demonstrated an outlook whereby Positive attitudes towards homosexuality were framed as Opinions (Pro), while Negative attitudes (Con) were framed as "Facts". This, again, deviates from the view whereby changes in country function as simply additives or combinations of perspectives. Searches using English in Morocco (fig. 15) resulted in the highest number of "Factual" Negative results in comparison to "Factual" Neutral results (21 Con vs. 6 Neutral). It would be all too easy to jump to a conclusion whereby these English language results have been passed through a filter of Moroccan conservativism. However, fig. 17 shows that, when using Arabic, the number of "Factual" Neutral results outnumbered the

“Factual” Con (15 vs. 13). Therefore, the context of searching using English in Morocco cannot be predicted by taking the results of searching using Arabic in Morocco and applying certain English search characteristics or vice versa by taking results using English in the UK and adding any perceived effects of a Moroccan search environment. These characteristics exist in a fluid and complex relationship with one another in which causal links are difficult to pinpoint. A key point also has to be reiterated, these results cannot be taken as representative, given that results continually change throughout time. Some of these changes will be outlined later on in this chapter, however, the principal conclusion to draw from these results is that there is a greater difference across contexts than can be accounted for by Google’s legally required censoring, shown earlier in their transparency report. It is clear that different contexts produce differently ranked results.

4.2 Comparisons of Language Differences

Users cannot search without a language, without a location, or outside of time. The examples of the last section show that there is a complex interplay between these classifications. For example, when the geographical domain was altered, there were few major changes to the search results, rather, it was in the interplay between location and other factors where its significance became apparent. In some contexts, location had much more of an impact than others. It is misleading to make generalisations that domain “x” has “y” effect on results, as many correlations did not hold across the categories; no criterion is neatly deterministic, instead, each category exists only in interrelation with others. Each search exists in its own network of associations whereby some criteria are emphasised more than others, at

different times, which leads to unexpected outcomes. However, the clearest separations were between language use, due to Google prioritisation of pages written in the language that has been searched.

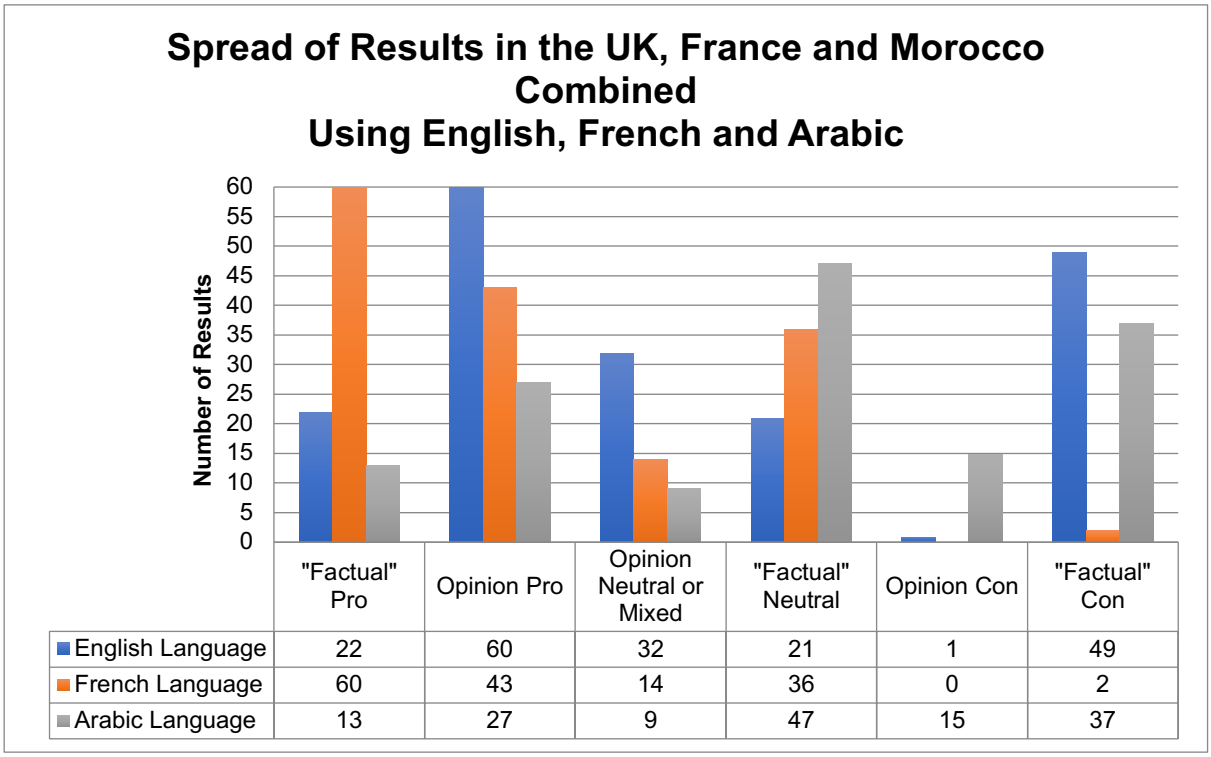
Fig. 18 shows the combined results of each country (UK, France, and Morocco) divided by language (English, French, and Arabic). There are some strong correlations that were indicated when discussing geographic domains. Also, although the effect of domain was demonstrated, comparing these combined results with the ones shown in fig. 19, which shows the results of the official language in each country, there is a similar pattern. Although domain affected the search results, language fixed the results more significantly. The following generalisations can be made about each language's results concerning homosexuality: in this study, searching using English provided Positive attitudes based around Opinion and Negative ones framed as "Facts"; searching using French provided an almost totally Positive perspective regarding homosexuality (in particular, searching using French provided a much higher number of "Factual" Positive results than either of the other two languages); searching in Arabic provided mostly "Factual" Neutral results with a high number of "Factual" Negative pages. It should also be noted that these generalisations still hold when compared to the results of fig. 19, that show the results of searching using the official language in each country. These results provide an indication of each context's epistemic landscape. Not only were there differences in sentiment, for example searching in French provides a much more Positive set of results than searching in Arabic, but the modes of discourse presented in the webpages provides an indication of what kind of an authority Google Search appears to be in each language. Due to the kinds of results given, in some contexts searching Google might seem like consulting a formally organised

source of knowledge, such as an encyclopaedia; in other contexts that provide a greater number of Opinion-based results, searching Google might be more akin to an online forum or radio talk show.

4.3 Mode of Discourse

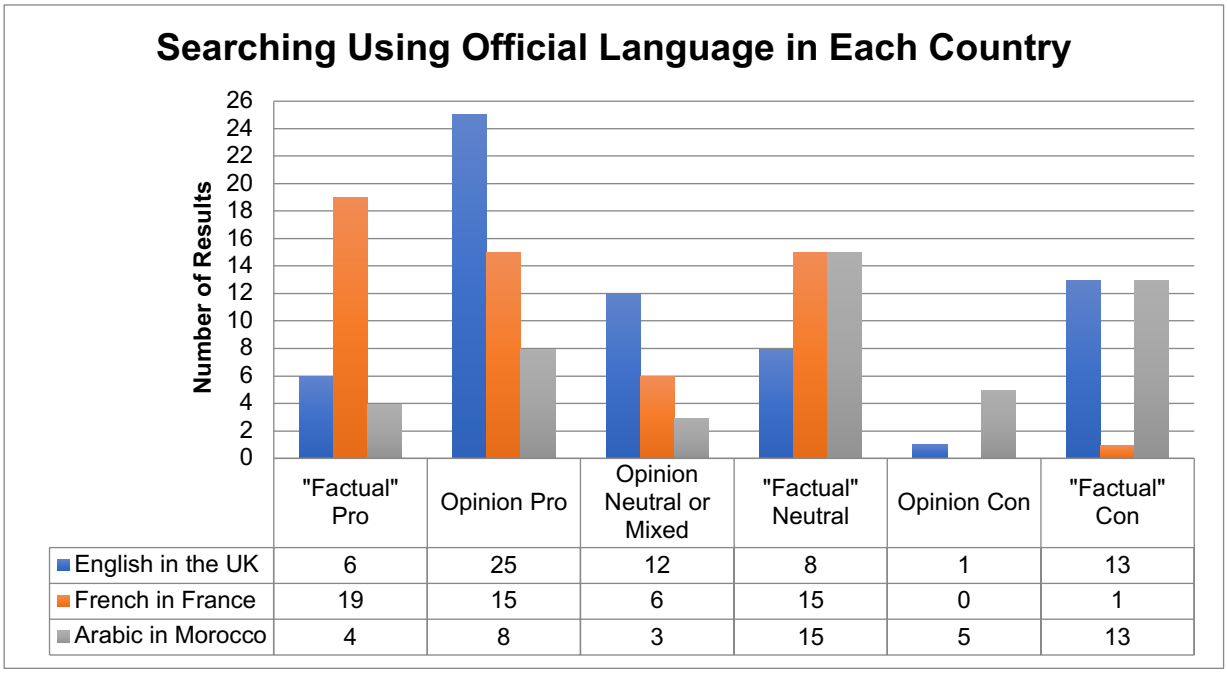
Shifting the focus away from whether results suggest a Positive or Negative attitude towards a topic such as homosexuality, we can also observe the discursive mode in which these pages were presented. Searching in different languages produced variation in regard to tone, authority, and perspective, which can be generalised by dividing results between pages that present information in a “Factual” or Opinion based representation. These two categories do not relate to how convincing or verifiable the arguments made were; instead, they are descriptions of the genre, format, or discourse of the webpage. Fig. 18 displays the combined data of each country, divided into languages used. The results shown in fig. 18 show that French and Arabic searches produced more “Factually” framed results whereas English had an almost equal split between these types and Opinion-based results. However, focusing directly on the data of each language in its official national domain provides a different insight (English in the UK, French in France and Arabic in Morocco, see figure 19). Although the results for French and Arabic stay in similar proportions, English appeared to be more affected by geographical domain. Using English in the UK produced more results in the mode of Opinion than when using English in other domains, where “Factual” pages were more common. In addition, by omitting the “Mixed or Neutral” results, in order to solely show the results that were directly

Positive or Negative (for comparison, see figs. 20 and 22, and 21 and 23, respectively) the differences become less significant.



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
English Language	22	60	32	21	1	49
French Language	60	43	14	36	0	2
Arabic Language	13	27	9	47	15	37

Figure 18. A graphical representation of the results for each location combined and divided by language.



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
English in the UK	6	25	12	8	1	13
French in France	19	15	6	15	0	1
Arabic in Morocco	4	8	3	15	5	13

Figure 19. A graphical representation of each language used in the specific country in which they are the official language.

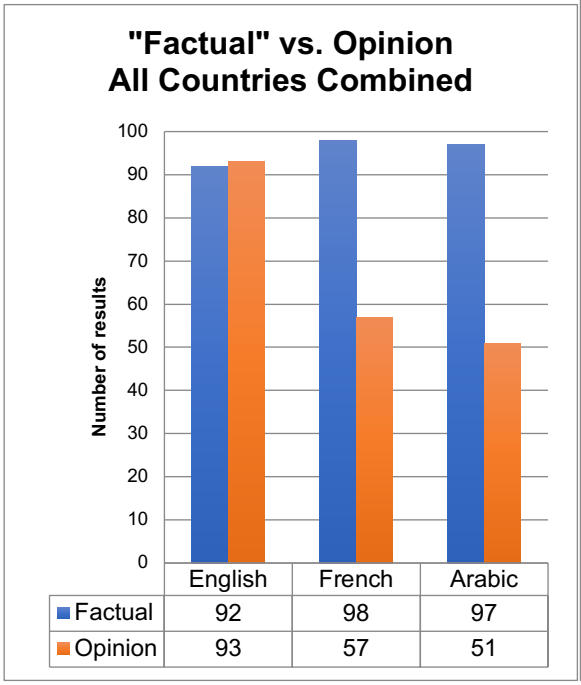


Figure 20. A comparison between "Factual" and Opinion pages, using each language (all countries combined).

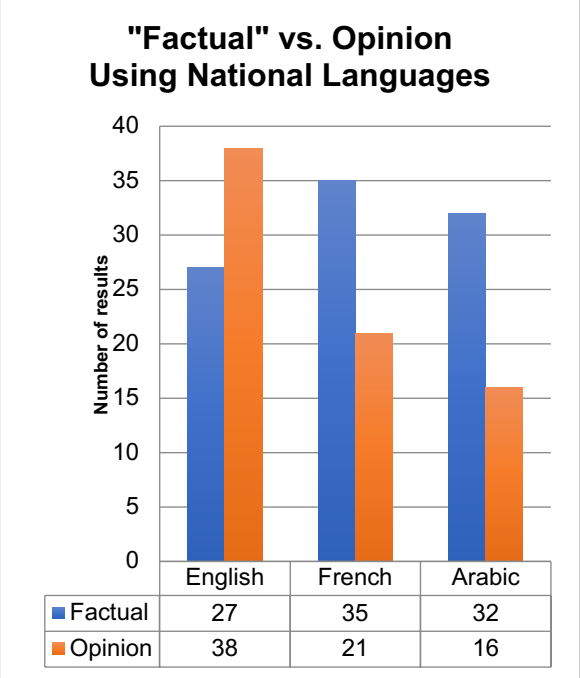


Figure 21. A comparison between "Factual" and Opinion pages, using only the national language for each country.

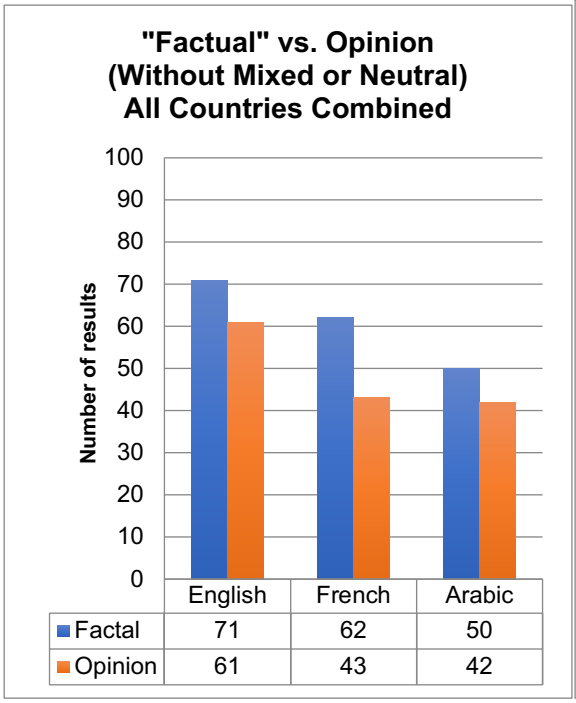


Figure 22. A comparison between "Factual" and Opinion pages without Mixed or Neutral results, for each language (all countries combined).

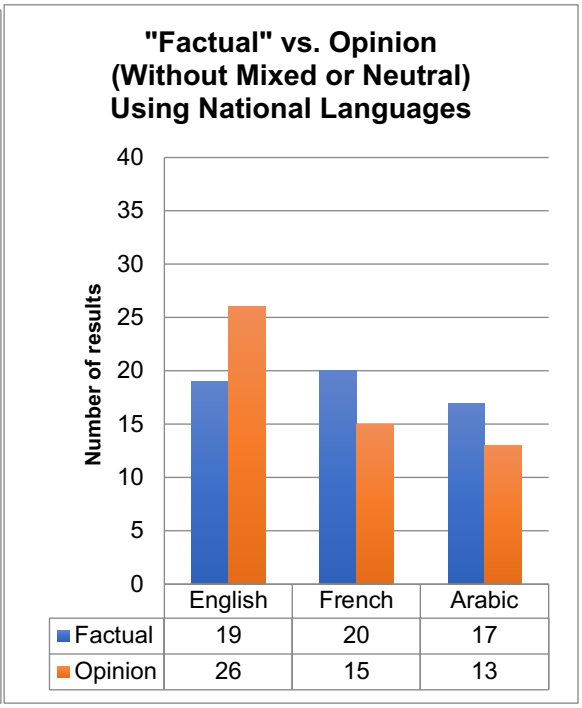


Figure 23. A comparison between "Factual" and Opinion pages without Mixed or Neutral results, using the national language for each country.

The *perception* of impartiality can develop from two separate modes of discourse: either as a collection of a range of different viewpoints (categorised here as Opinion Neutral or Mixed and “Factual” Neutral) or a single viewpoint that addresses its audience in a “Factual” tone, which could take a Positive, Negative or Neutral stance. Which kinds of voices are highlighted and which are hidden is a complex question, but the data suggest that Google’s promotion of Wikipedia, medical FAQs, moral claims backed by biblical scripture and other “Factual” sources outweighs other modes of communication. This should be particularly concerning when such “Factual” webpages align more often with a particular viewpoint, as they have in this case with a distinctly anti-homosexual sentiment.

So far, observing the consequential factors in each search has provided an indication of how particular characteristics, or contexts, shape results. Geographical difference had less of an effect than altering the language used. In particular, geographical difference produced noteworthy effects only via particular combinations of countries, rather than directly. Changing language consistently had an effect on results, whereas changing geographical location had different levels of influence depending on the language. The next section addresses the *kinds* of words used: how much impact does word choice, vocabulary, and expression have on results; essentially, does the way in which individuals phrase a query determine the kinds of results they receive?

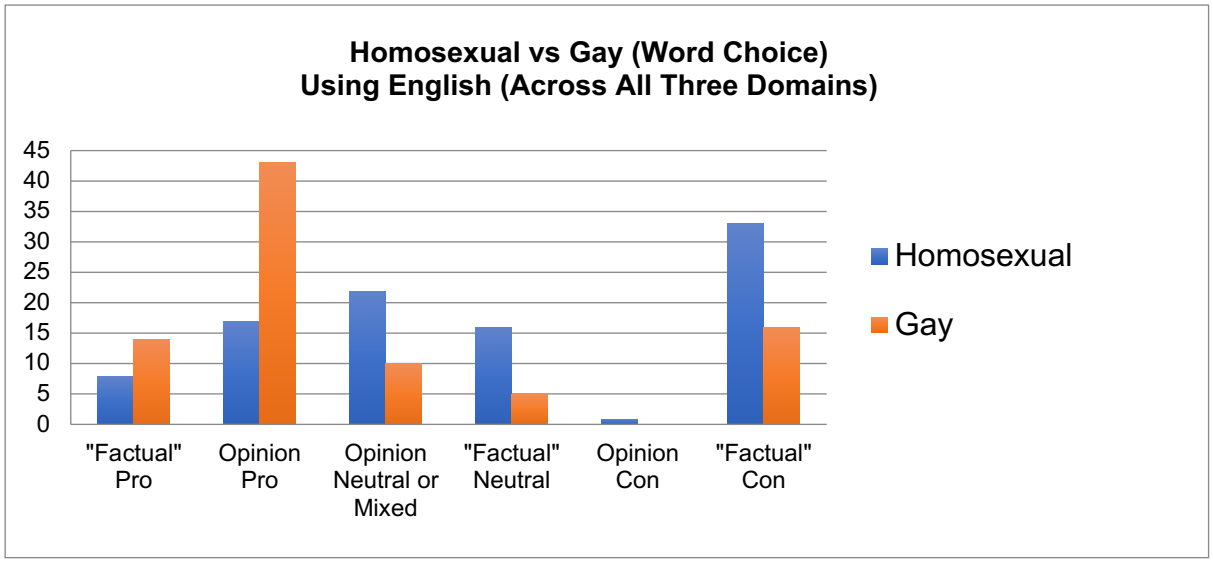
4.3.1 Word Choice and Phrasing

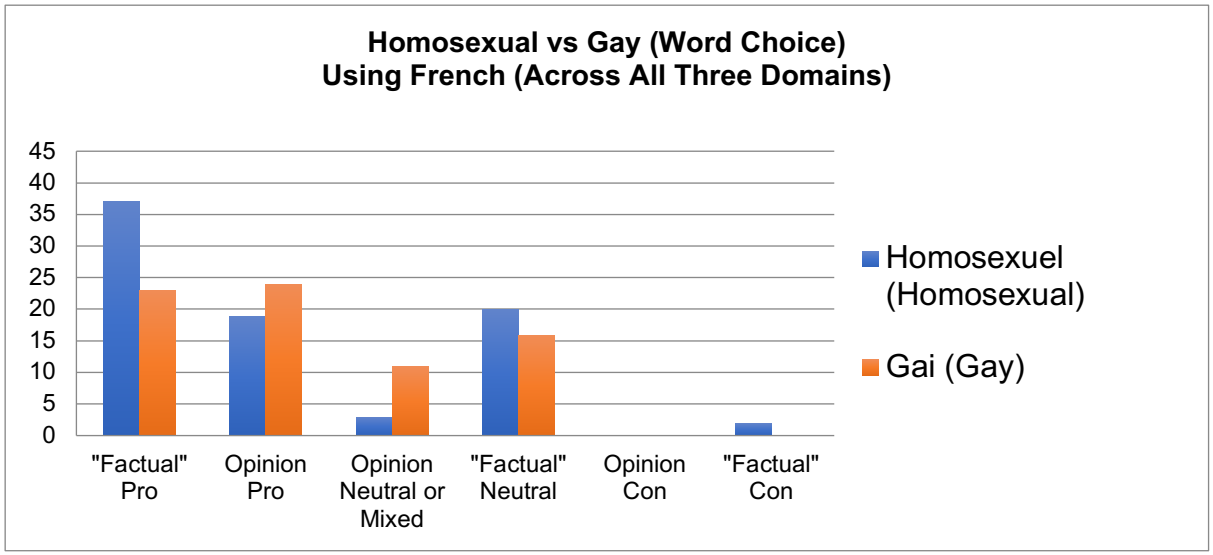
In the context of this study, two kinds of alterations were made for comparison. First, the experiment tested a linguistic variation of each query, searching each query with the word [homosexual] and the word [gay], including their translated equivalents in each language. This variable was referred to as *word choice* and aimed to represent formal and informal vocabularies. Second, a linguistic variation was introduced for the queries phrased as questions, substituting and comparing queries phrased with the word [wrong] or the word [good]. This, again, was replicated for each of the three languages and aimed to highlight the *phrasing* of queries that carried either a Positive or Negative sentiment.

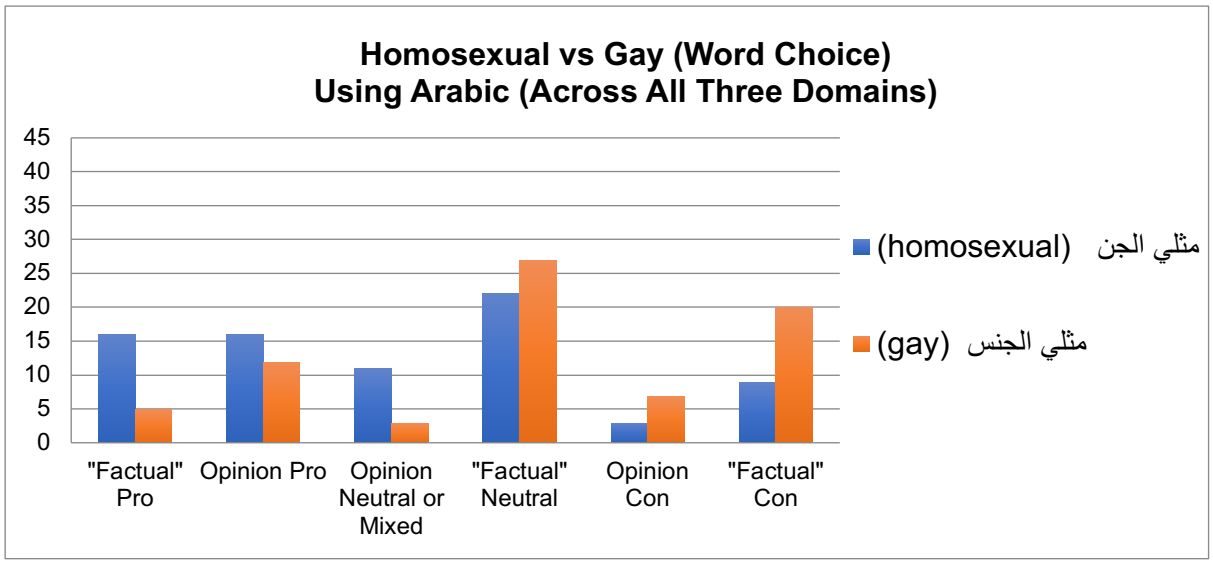
4.3.2 Word Choice – Gay vs. Homosexual

The following graphs (figs. 24-26) represent the results of queries that were identical apart from the use of either [gay] or [homosexual]. Each set represents an open-ended search (simply the single word without context, i.e. [homosexual] or [gay]), one Negatively phrased search (i.e. [Is being homosexual wrong?] and [Is being gay wrong?]), and one Positively phrased search (i.e. [Is being homosexual good?] and [Is being gay good?]). The difference in results was clear, using [gay] produced more Positive and Opinion based pages, while [homosexual] produced more Negative and “Factually” framed results. Using English in all three domains (see fig. 24) provides a clear split: using “gay” in a query, such as [Is being gay wrong?], provides an overwhelming number of Positive Opinions, while search queries using the word

“homosexual”, such as [Is being homosexual wrong?], resulted in a dominance of Negative “Factual” results.







Figures 24-26. Graphical representations of word choice between [homosexual] and [gay] and respective variants in each language, all locations combined.

This shows the way in which two communities have grouped around these two terms: in the results analysed, “gay” was deployed in a way that draws on personal and Positive perspectives; “homosexual” was deployed more commonly in Negative religious or medical contexts, in which same-sex relationships are described as a sin or an illness. The most noteworthy aspect of these results is that, rather than synonyms being ignored, there is a significant crossover between the wording of search terms and the wording of results. Google Search uses synonyms for each term but, in these examples, appeared to draw on results that are *representative* of the main association of each term; i.e. the word [homosexual] in queries triggered webpages with Negative (Con) and “Factual” content, while using the word [gay] in queries triggered Opinion based Positive results. The results for [homosexual] often included pages that repeatedly relied on the word “gay”, rather than the word “homosexual”; however, these were uncharacteristically “Factual” and Negative compared to any of the results that mainly relied on the word “gay”, rather than “homosexual”, which were provided by queries that used the word [gay]. By the same token, when webpages using the word “homosexual” appeared as the results for a search of the term [gay], they were regularly Positive and Opinion based. This is in contrast to the kinds of pages using the word “homosexual” usually returned for a query using the word [homosexual]. For example, fig. 27 titled “What Does the Bible Say About Homosexuality” shows a webpage that is returned as a first-page result for the query [Is being gay wrong?] but not [Is being homosexual wrong?], even though that page uses the word “homosexual” thirteen times and the word “gay” only eight times. The page is one of the only religious results that is Pro and written in an Opinion discourse, see the final paragraph inside the red box.

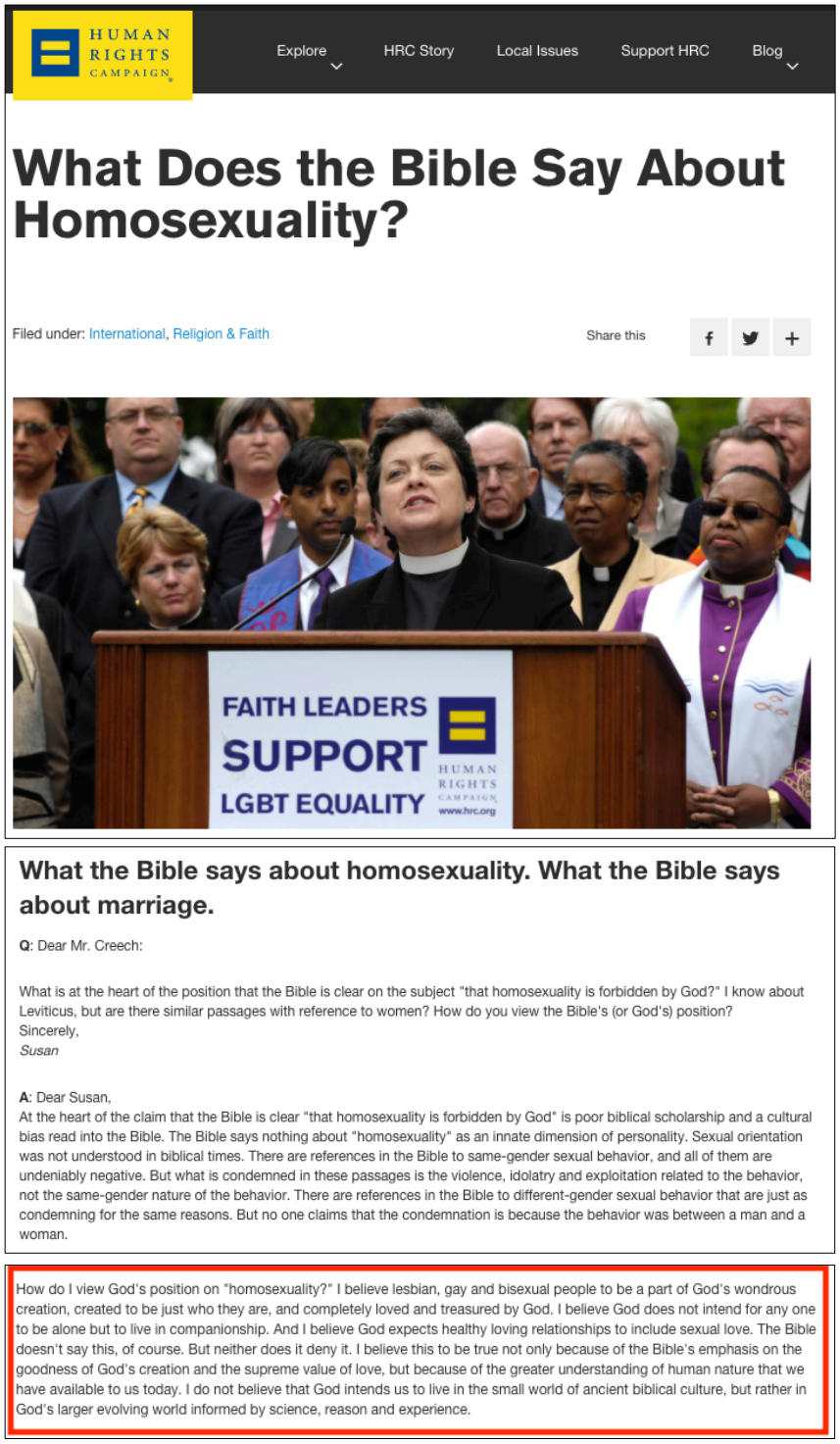


Figure 27. “What Does the Bible Say About Homosexuality?” (Chreech). A webpage result for the query [is being gay wrong?] and is rated as Pro Opinion. This webpage is not ranked as a result in the first page for [is being homosexual wrong?] even though the number of times this page uses the word “homosexual” is almost double the number of times the word “gay” is used. Accessed 22/08/2017.

Comparing this result to the graphical representation of the overall discourse in fig. 24 shows that the result fits with the Pro Opinion context that searching using the word [gay] in a query produces. That fig. 27 is not being given as a result when searching [Is being homosexual wrong?], even though it uses the word “homosexual” far more frequently, shows that the results have aimed to match the predicted underlying sentiment of a user, indicated by the phrasing of a query, rather than simply matching word usage.

If we return to the previous discussion on relevance, there is a clear sense that the word usage of the user formulating a query is being interpreted for an underlying sentiment behind their choice of words. Such a finding demonstrates the way in which Google’s algorithms are designed to provide *relevant* results. This notion of relevance sits between Group A’s notion of objectivity and Group B’s personalisation; Google’s algorithms and machine learning techniques have inscribed the social values of the words used into the way it ranks a set of results. If on average, users searching using the word [gay] – instead of the word [homosexual] – are more Positive towards same-sex relations, then a more Positive set of results is likely to be more *relevant* to that user, regardless of the actual vocabulary used in the webpage. The conclusion of chapter one is relevant here: search queries are different from questions in that they are attempts to iterate the language of a result that the user hopes to find. Early search engines searched for exact linguistic matches in order to fulfil the needs of a user. The results of the above study indicate that Google have taken this process a step further. The use of Wittgenstein’s quotation “*The limits of my language* mean the limits of my word” (68 emphasis in original) in chapter one (1.3), reinforced the technological affordances of

search engines before the default utilisation of synonyms.[7] However, Google's technological innovations have intensified, rather than nullified, the prison-house of language. The above results suggest that in their drive towards *relevance,* Google's results are tailored to the associations of particular words and the sentiments that most commonly accompany their usage.

Currently, the discourse regarding the personalisation of results centres around the potential aggregation of data gathered about specific individuals, from their cookies, search histories and information available from cross-platform log-in accounts, for example using information from an individual's Gmail account to tailor Google Search results. Zimmer's argument above (section 2.2) regarding dataveillance takes this position, arguing that "Google, like most for-profit search engine providers, is financially motivated to collect as much information as possible about each user: receiving personalised search results might contribute to a user's allegiance to a particular search engine service" (92). In addition, many of the Group B theorists above orient their position with regards to some kind of individual-specific profiling or, to use legal scholar Daniel Solove's term: a reliance on "digital dossiers" (2). However, I argue that, based on the result of this experiment, the way in which cultural values and attitudes have been incorporated into algorithmic decision-making is far more insidious than data-driven personalisation for specific individuals. Using a VPN meant that personalisation was not a factor in the results. However, the variation in the kinds of results for each query demonstrates that results are being tailored depending on the kinds of word choice and phrasing used. If, through individual-specific data-collection, a user was deemed to have a particular attitude and, consequently, their search results were being shaped to fit those attitudes,

[7] See footnote 9 in chapter one section 1.3 regarding Google's introduction of synonyms as a default.

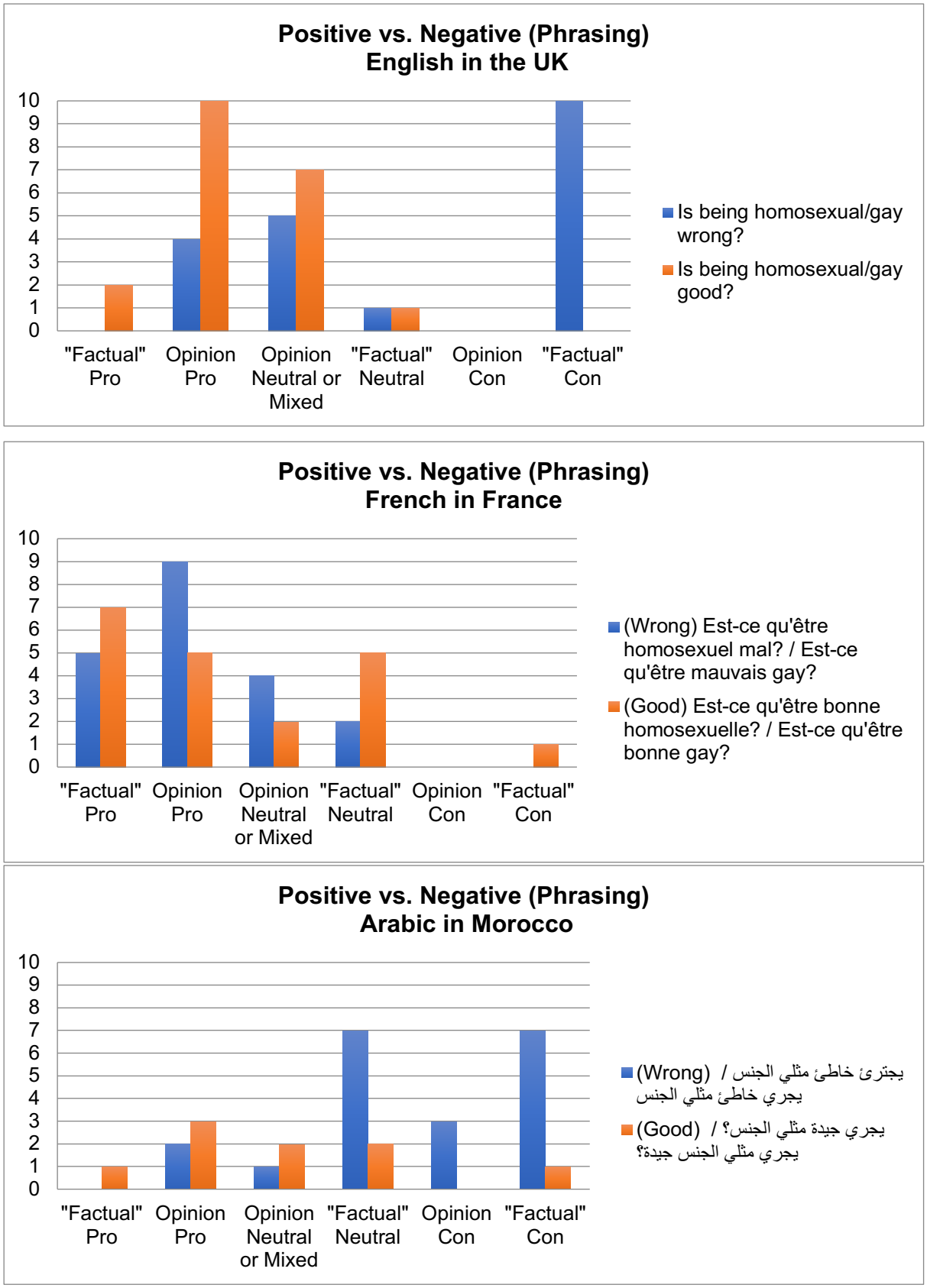
there would be clear ways of breaking out of such personalisation. The user might perform a search on an IP address not related to them, a computer of a friend or at a library, and they would see how their results were different. However, if the results are *instead* shaped by linguistic modes of discourse, particular vocabularies and patterns of phrasing, which seems to be the case, users are stuck in a less-visible feedback loop whereby their sentiments are continually reinforced.

Other languages showed different kinds of associations with each word; each word that was varied represented different attitudes and therefore the results differed greatly. In French, using [homosexuel] rather than [gai] produced results that were much more "Factual" and Positive (see fig. 25). However, like the English example, Google's use of synonyms means that each word appeared in the other's results (pages which use the word "gai" returned in the results for queries using the word [homosexuel]) but only pages that represented the dominant characteristics of the other word (pages that were predominantly Positive and "Factual" but still mainly use the word "gai"). Searching in Arabic also produced differences in results between queries that used the word homosexual or the word gay in Arabic. [مثلي الجن] ([homosexual]) provided more Positive results whereas [مثلي الجنس] ([gay]) provided mainly "Factual" Neutral and "Factual" Negative results (see fig. 26).

4.3.3 Word Choice - Phrasing as Positive or Negative

This final section of analysis focuses on queries that were divided along the lines of Positively or Negatively phrased questions. Figs. 28-30 show the results of these pre-framed questions and demonstrates a clear instance of Google results that are tailored towards a particular kind of relevance. Phrasing a query by asking [Is being

homosexual wrong?] or [Is being gay wrong?] provided results that supported the implicitly Negative phrasing of the statement (see fig. 28). Phrasing a query using the word [good] instead of [wrong] resulted in a large number of Opinion Pro results; in addition, the second most dominant kind of results were Opinion Neutral or Mixed, which shows that even the results that were not explicitly Pro homosexuality were framed as Opinions. Fig. 29 shows that there was no such correlation when searching using French in France; the overall language and domain biases towards presenting homosexuality in a Positive light remained dominant. For reasons mentioned earlier, the same searches conducted using Arabic in Morocco reveal a complex selection of results. Although the graph shows that the results were mainly Negative, only text-based results concerning homosexuality were included. When searching using a Positive phrasing [الجنس مثلي جيدة يجري؟] and [يجري مثلي الجنس جيدة؟] ([Is being homosexual good?] and [Is being gay good?], respectively) the results returned same-sex pornographic video content (fig. 30). It is difficult to know whether this is due to the association of this particular translation or whether these results make contextual sense for that query. Regardless, these results were not counted as Positive Opinions because they did not directly represent an attitude towards homosexuality, comparable to the other kinds of results.



Figures 28-30. Graphical representation of positive and negative query phrasing of languages searched where they are national languages.

Focusing on the scale of an individual user, if an individual were searching queries about homosexuality, to get a sense of what their surrounding community felt towards the issue, hardcore homosexual pornography could either be perceived as a liberating vision of acceptance or a way of contextualising homosexual experience as deviant. Homosexual pornography was common in the results of searches using Arabic in the UK and France; there was only one pornographic result from [يجري مثلي الجنس] ([Is being gay good?]) searched in Morocco, i.e. homosexual pornography was more common in the results of queries using Arabic *outside* Morocco. When the query [الجنس مثلي جيدة يجري؟] ([Is being homosexual good?]), a search that provided mostly hardcore gay pornography in the UK and France, was searched in Morocco, Google returned no results and instead returned the message "Your search did not match any documents. Suggestions: [none] Try different keywords", which is a clear sign that the results had been directly censored, because searching this same query in the UK or France produced results, meaning that there are documents in Google's index that match that query. Google is open about their adherence to the local laws of specific countries, which force them to alter results (homosexuality is, after all, illegal in Morocco).[8] However, there is a disconnect between the Positive results that can be found both by searching using Arabic in other countries and searching using other languages in Morocco. The fact that a whole set of search results has been banned, or "did not match any documents", would likely suggest a mixed perspective to the user. That the results for [Is being homosexual good?] are banned but other searches related to homosexuality are not and, in fact, actually present some Positive results,

[8] All legal removals can be viewed at Google's "transparency report" section of their website. However, it should be noted that the Moroccan section claims that they have only ever received two requests for removal in Morocco. Such a claim does not fit with the missing page. However, the report does not detail any unofficial agreements to hide results. This may well be evidence of Google censoring results far more than they claim.

foregrounds the complexity of the issue at hand, within that context, less through the content of specific pages and more through the overall search landscape.

The fact that altering the word usage can provide such a spectrum of results, from hardcore pornography to complete censorship of results, shows the malleable nature of searching, which shapes around different contexts and different users. These search characteristics cannot be treated as simplistically as Morocco equals conservative or Arabic equals constricting. Searches in Arabic were the only ones that resulted in such graphic demonstration of homosexuality: no other language produced pornography from the queries searched. Simply using a Positive phrasing in Arabic produced homosexual pornography, an indication that such results may be very *relevant* within a culture that represses such activities. That searching in Arabic provides *both* Negative views towards homosexuality (that might be dominant within the offline culture of Morocco) and very direct depictions of pornography provides a complex picture. Both sets of results derive from similar contexts; the pornography may be symptomatic of the tabooed nature of homosexuality that is also the source of highly charged Negative webpage results. However, these two contrasting attitudes are the results of very similar search queries, with only minor changes in vocabulary. Consequently, although many of the webpage results overlap, subtle and potentially unintentional word choices keep searchers in their own context of relevance.

5.0 Unimaginable Communities

With the above study in mind, I want to draw from Benedict Anderson to suggest that these changing contexts can be described as *unimaginable communities*. The meaning of Anderson's original term "imagined communities", describes the interrelations of print capitalism with the rise of nationalism. Anderson argues that imagined communities developed due to the movement away from Latin and the rise of new reading publics focused around localised vernaculars which emphasised linguistic and geographical boundaries. Such distinctions created communities of people who, although they would never meet, shared some commonality, out of which the modern conception of the nation developed. In a digital context, such issues of nations, language, and publics are still deeply important. The identities of Anderson's individuals developed in concert with a wider sense of community. Today, we still take part in contextualising our thoughts and actions within a community. These actions are not necessarily taken with communities *in mind*, but the locations, languages, dialect and tone, determine each individual's algorithmic milieu. Algorithmic governmentality and the instability of changing results means that although Anderson's framework is in some ways a useful metaphor, the kinds of communities he discusses, even imagined ones, do not map onto our present experiences. Context, or community, is a vital part of Google Search but it is not just that these communities are harder to imagine than the nineteenth-century equivalents. These communities are fundamentally *unimaginable*; the formations and calculations that occur for each Google search are so complex that individuals exist in an unimaginable arrangement. The heterogeneous elements change throughout

time but they constantly produce patterns and correlations that make repeated actions predictable and relevant.

In their 2013 book *The New Digital Age: Reshaping the Future of People, Nations and Business*, Eric Schmidt (Executive Chairman for Google) and Jared Cohen (Director of Google Ideas) present their vision for a global digital future. They share a utopic optimism that "We will increasingly reach, and relate to, people far beyond our own borders and language groups, sharing ideas, doing business and building genuine relationships" (6). They argue that this new digital age (by which they often mean Google) "will usher in an era of critical thinking in societies around the world that before had been culturally isolated" (34). New technology, they say, will break down barriers and help us to live better lives as more informed world citizens:

> People who try to perpetuate myths about religion, culture, ethnicity or anything else will struggle to keep their narratives afloat amid a sea of newly informed listeners. With more data, everyone gains a better frame of reference. A Malawian witch doctor might find his community suddenly hostile if enough people find and believe information online that contradicts his authority. Young people in Yemen might confront their tribal elders over the traditional practice of child brides if they determine that the broad consensus of online voices is against it, and thus it reflects poorly upon them personally. Or followers of an Indian holy man might find a way to cross-reference his credentials on the Internet, abandoning him if it is revealed that he misled them. (35)

It is easy to be dismissive of the tone the authors use. They talk brazenly with their biases on their sleeves as if all knowledge works unilaterally from total ignorance to wholly developed thinking that aligns them with the attitudes of Group A thinkers, outlined above. They choose their examples carefully to play to their audience; however, the tone used conveys an attitude that, pre-internet, most cultural niches are simply examples of brainwashing and that technology will inevitably bring the truth. Regardless of the techno-deterministic optimism, their text is strewn with moments of doubt, fifty pages after the above idealistic quotation, they state that "Evidence shows that most internet users tend to stay within their own cultural spheres when online, less for reasons of censorship than because of shared language, common interest and convenience" (85). They do not indicate what evidence they have in mind, but the kinds of search result arrangements demonstrated above might indicate that their evidence is particular algorithmic arrangements of Google.

The results above demonstrate how complex the interrelations of community are and how variations within these characteristics create sub-communities or niches that experience the web differently. Even if Google were a company devoted to producing world citizens, rather than capital, how might this balancing of characteristics work? There is no doubt that the examples of Malawian witch doctors and Indian holy men are problematic. But in very real terms, how do we balance the need for information to be *relevant*, to make sense in our context, with a broader range of views and ideas? This is less a question to be answered, rather than one that is enacted in every Google search. Every set of queries presents a statement. The difficulty is that, as seen in the variations in phrasing, intention, language and geography, it is not even clear to users what kind of results they have received.

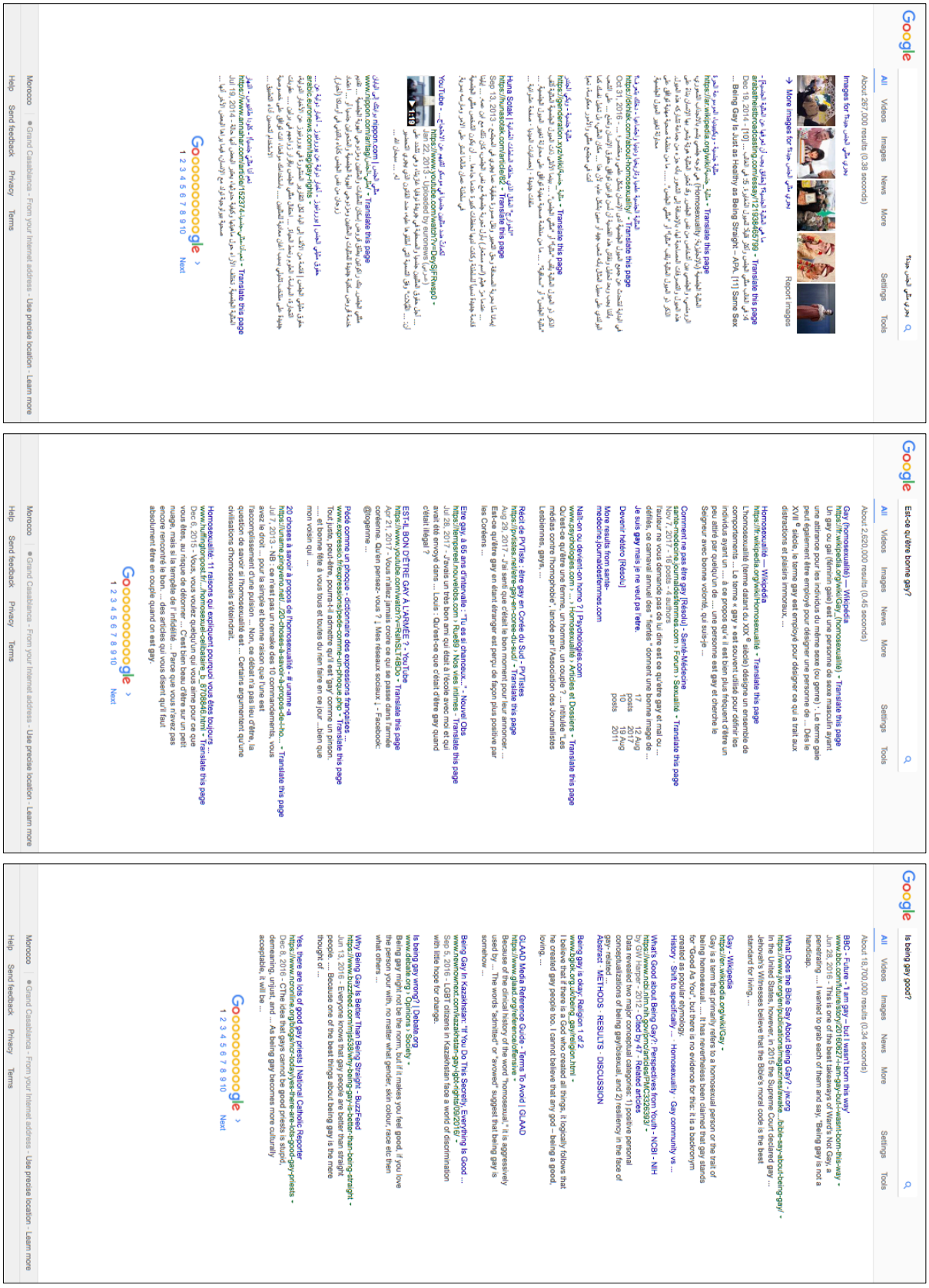
Users do not know whether their results are very different in kind to someone searching using a different language in their same country, perhaps even next door. It is unlikely that the average French Google user is aware that their results are notably favourable towards homosexuality. Or that the results of English speakers fluctuate more than other languages when they search from different geographical locations. We do not know this because to do so would mean understanding a much larger picture than is possible for the average user. Instead, we must rely on *unimaginable communities.* Users may be aware that Google results have developed from the success of other users, but who these community members are is a complex matter, so too is the metric of a *successful* search result. These communities cut across boundaries to form shifting assemblages of relevant criteria, therefore these communities are not difficult to imagine but impossible to imagine. The use of data has reinforced old distinctions that separate out contexts, but in doing so, have obliterated the fixed membership of those groups. There is a plurality of Googles, but the number that any user can access is restricted, not by filters, but by much older and more nebulous aspects of communication that shape experience, even in a digital landscape.

6.0 Repeating the Experiment 24 Months Later

The searches discussed above took place on 24 November 2015. On 22 November 2017, the previous searches were repeated to see if the results had changed in any substantial way. The results had changed; many of the different contexts had standardised to such a degree that comparing every example with those from the 2015 tests in full was deemed unnecessary. As outlined above, the nature of Google's algorithmically produced results, means that the particular changes noted, between 2015 and 2017, cannot represent generalisable evidence. Instead, highlighting some of the changes demonstrates that search results change throughout time, not simply in the pages returned to a user but also, the overall tone, attitudes and broader epistemic landscape presented. In light of chapter three's conclusions, regarding the increasing importance of RankBrain as the third most significant signal in search results, I posit that the increasing standardisation between different contexts is potentially a direct result of Google's neural network machine learning. As discussed in chapter three, RankBrain's core behaviour is rephrasing queries in order to measure related user behaviour and *learn* about the relationship between linguistic phrases, in order to better predict the user sentiment that lies behind a query's phrasing or vocabulary choices. RankBrain was originally designed to use the data regarding common queries to inform the ranking process for uncommon or unique queries. Therefore, using information about common search contexts, for which Google's algorithms have more data to draw from, to inform the ranking of less common search contexts would be a clear extension of RankBrain's original purpose.

6.1 Comparisons of Language Differences 24 Months Later

Differences in the actual webpage results for each language were still significant; this underlines the continuing importance of the logic of early search engines, outlined in chapter one, that Google is finding pages in which those words appear (see figs. 31-33).



Figures 31-33. Moroccan Search results for the query [Is being gay good?] using Arabic, French and English. Accessed 22/11/2017.

Although, as in 2015, synonyms *were* clearly being used to find results, the pages retrieved were all written in the language searched in. While Google does sometimes provide pages written in a different language to that which was used to search, with the option of translating it using Google Translate, this was not found to be the case for any of the queries tested. This shows that individuals searching using different languages are still receiving different kinds of content, regardless of any similarities of their underlying attitudes. However, there was a rise in particular institutions gaining prominence across all the varying contexts, two in particular: Wikipedia and jw.org, a Jehovah's Witnesses site run by the Watch Tower Bible and Tract Society of Pennsylvania. All the Wikipedia results were the language specific sites, each featuring different content; therefore, each language's Wikipedia page for the same topic is written separately and focuses on different topics. However, the Jehovah's Witnesses page that dominates so many results for different languages and domains is simply a series of translations of the same page (see fig. 34 for that page) and other related pages from JW.org (see fig. 35 for examples of languages and countries that feature the page represented in fig. 34, as well as other JW.org pages as highly ranked results).

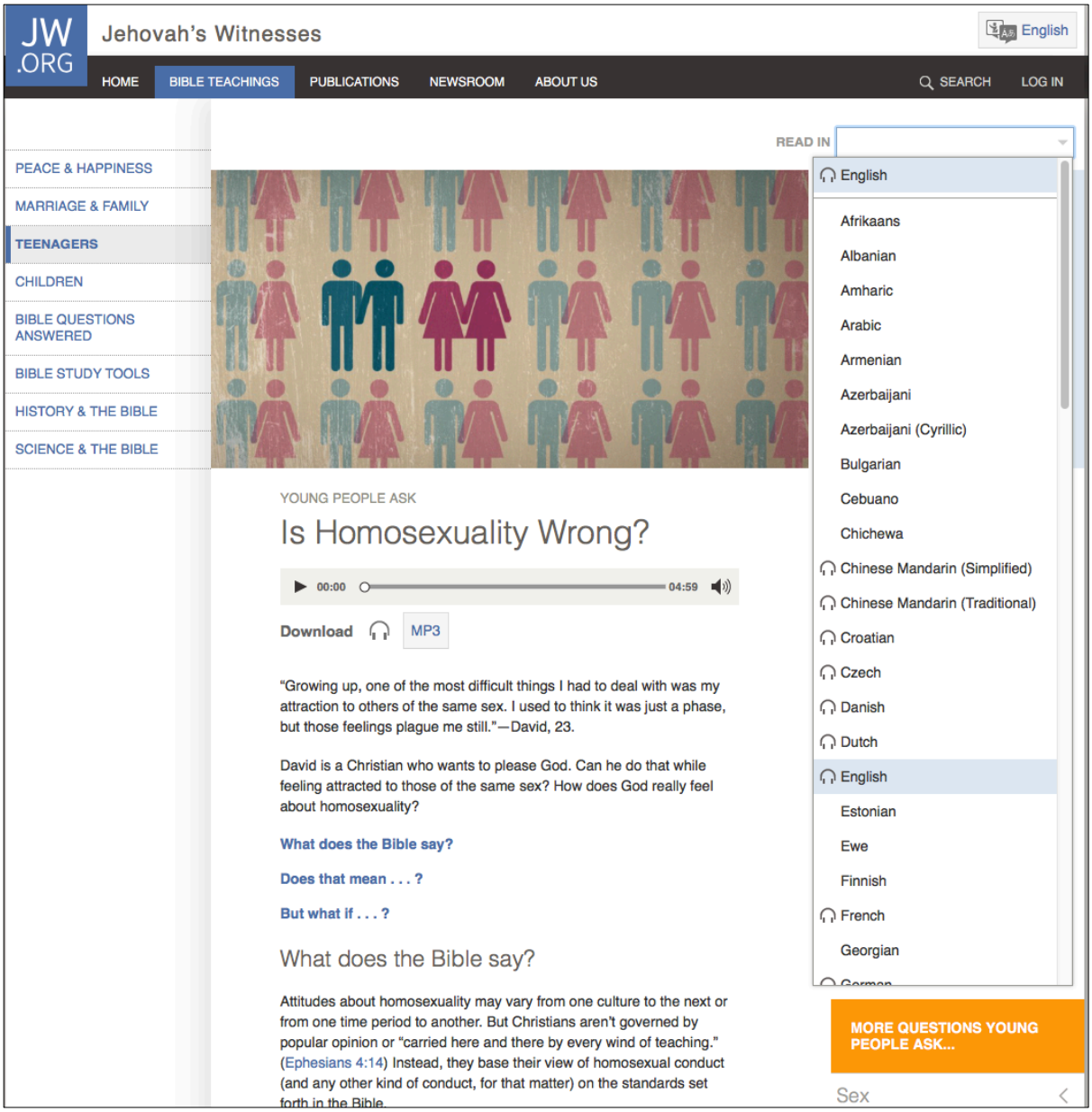


Figure 34. The Jehovah's Witnesses page "Is Homosexuality Wrong?" that is commonly provided by Google as the first result for a number of different languages and locations. The page is categorised as "Factual" Con. It bases its arguments by citing bible scripture throughout and framing such discourse as an evidence-based absolute rather than in the mode of Opinion-based argument. Although the page does not promote homophobia, the page has a clear attitude that homosexual acts are forbidden by God and those practising homosexuality should change their behaviour (see Jehovah's Witnesses). Accessed 22/11/2017.

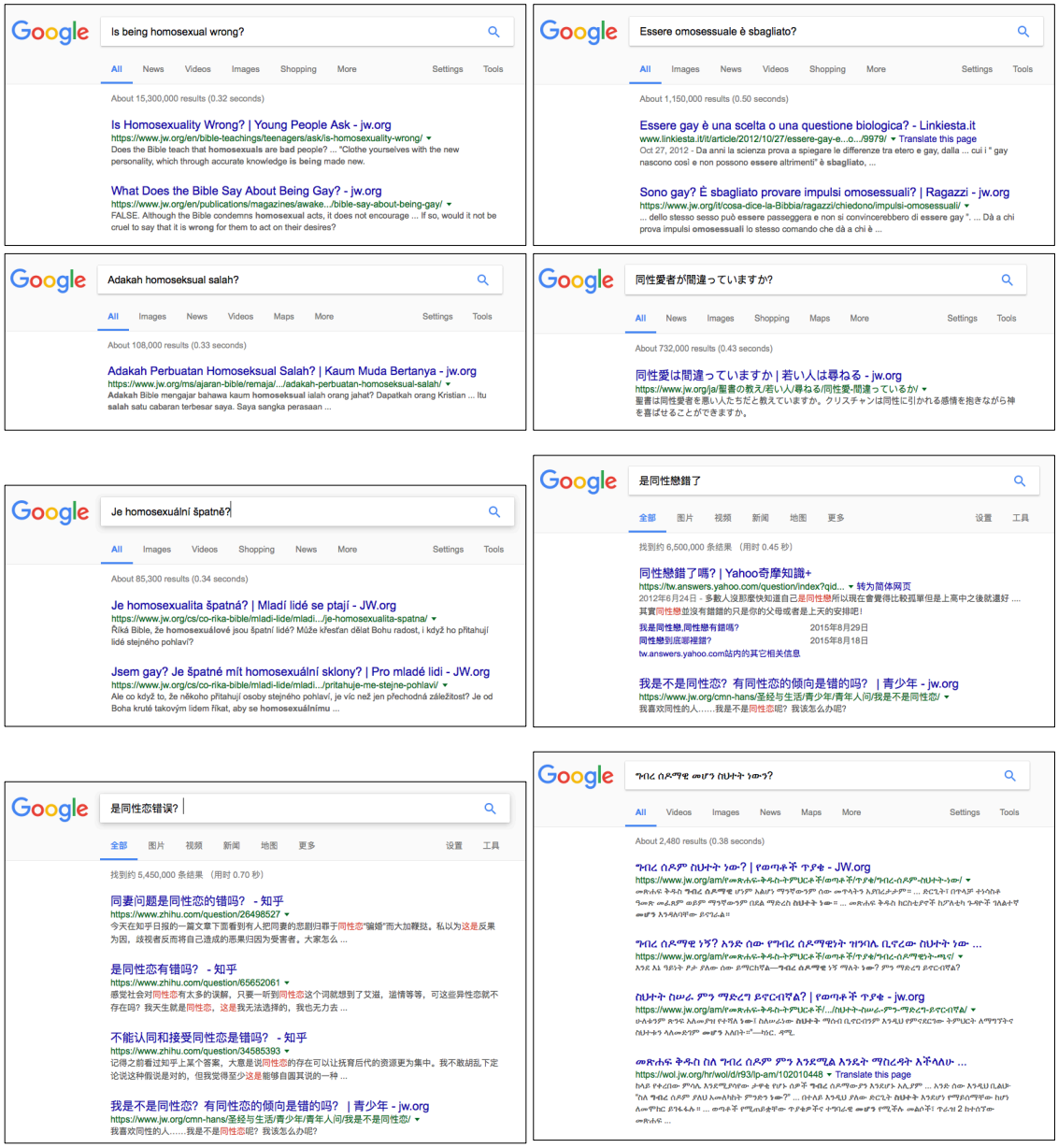


Figure 35. Search for the query [Is being homosexual wrong?] using different languages in various countries showing the Google results for "Is Homosexuality Wrong? | Young People Ask - JW.org" and other JW.org pages. Examples are as follows: English in the UK (1st result), Italian in Italy (2nd result), Malay in Malaysia (1st result), Japanese in Japan (1st result), Czech in the Czech Republic (1st and 2nd results), Chinese Traditional in Hong Kong (2nd result), Chinese Simplified in Hong Kong (4th result), Amharic in Ethiopia (1st-4th results). Accessed 22/11/2017.

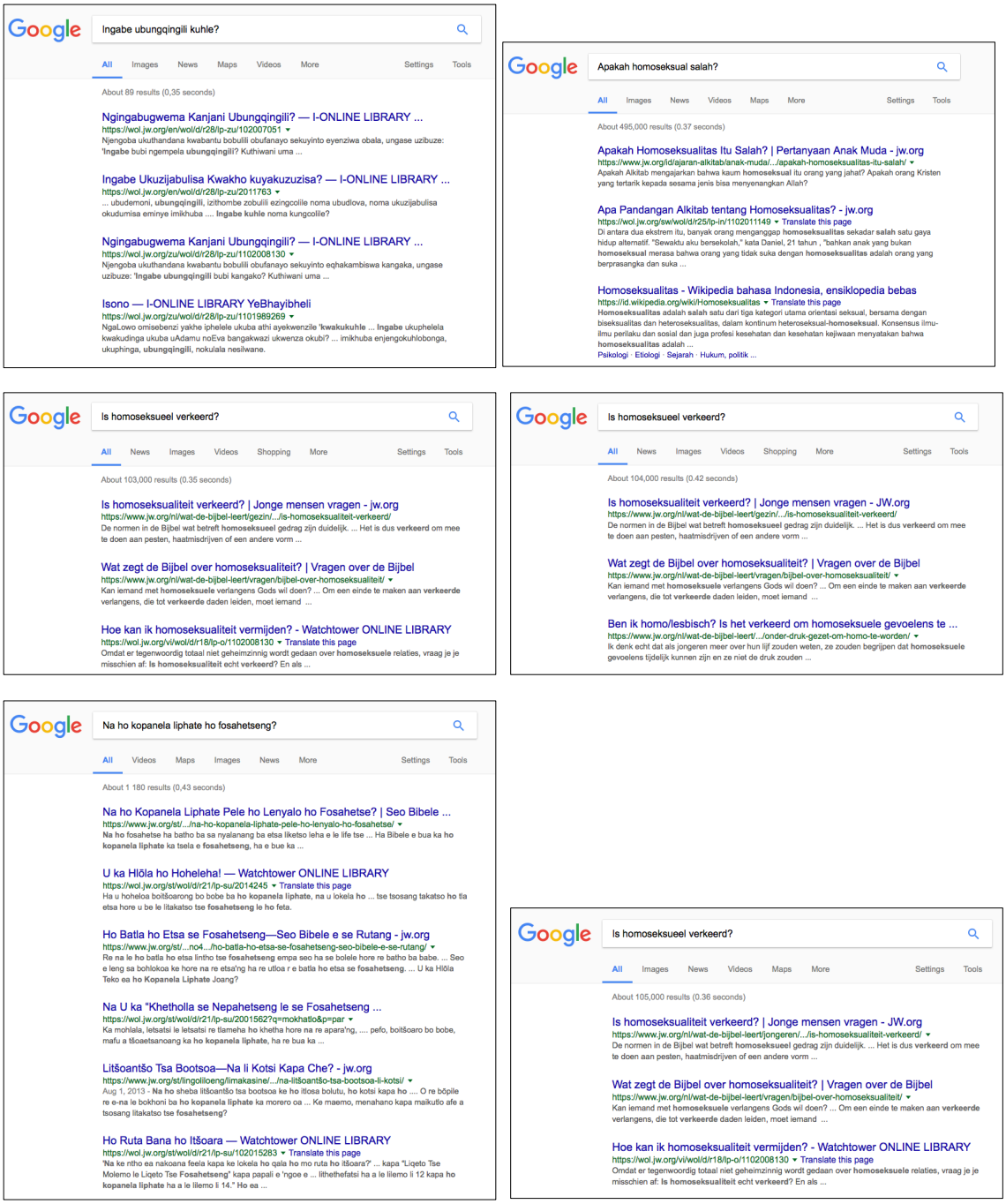


Figure 35 (continued). Search for the query [Is being homosexual wrong?] using different languages in various countries. Examples are as follows: Zulu in South Africa (1st-4th results), Indonesian in Indonesia (1st and 2nd results), Afrikaans in Australia (1st-3rd results), Afrikaans in the UK (1st-3rd results), Southern Sotho in South Africa (1st-6th results), Afrikaans in South Korea (1st-3rd results). Accessed 22/11/2017.

In the 2015 set of results there was no single overriding webpage result across different contexts, let alone one so prominent as to recurrently be returned as the first or second result. In many ways, the 2015 results showed a web balkanised division between languages and country domains. The multiple translations available for the Jehovah's Witnesses' page (fig. 34) does show that there is a mission to spread that particular message to as many different people as possible. However, drawing from an understanding of how search engines work, which was outlined in chapter one, I find it unlikely that each translation of the same page became the first result in so many different contexts, some of which have much larger, more varied, populations of users than others. Instead, I hypothesise that RankBrain is causing a normalising between these contexts. RankBrain's original task was rephrasing unique queries into more common previously searched ones, by predicting that the two queries had a high level of similarity, as determined by aggregated user actions and contextually similar linguistic patterns. As highlighted in chapter three, RankBrain became the third most important signal for ranking results between 2015 and 2017. The results demonstrated in fig. 35 suggest that RankBrain is using the results of a query from a context where it has been searched a higher number of times, and consequently has more data from which to draw from, in this case [Is being homosexual wrong?], and using it to inform contexts in which an equivalent query has been searched fewer times. This would lead to this single page dominating in so many different contexts and particularly in niche contexts, such as using Afrikaans in South Korea (see fig. 35).

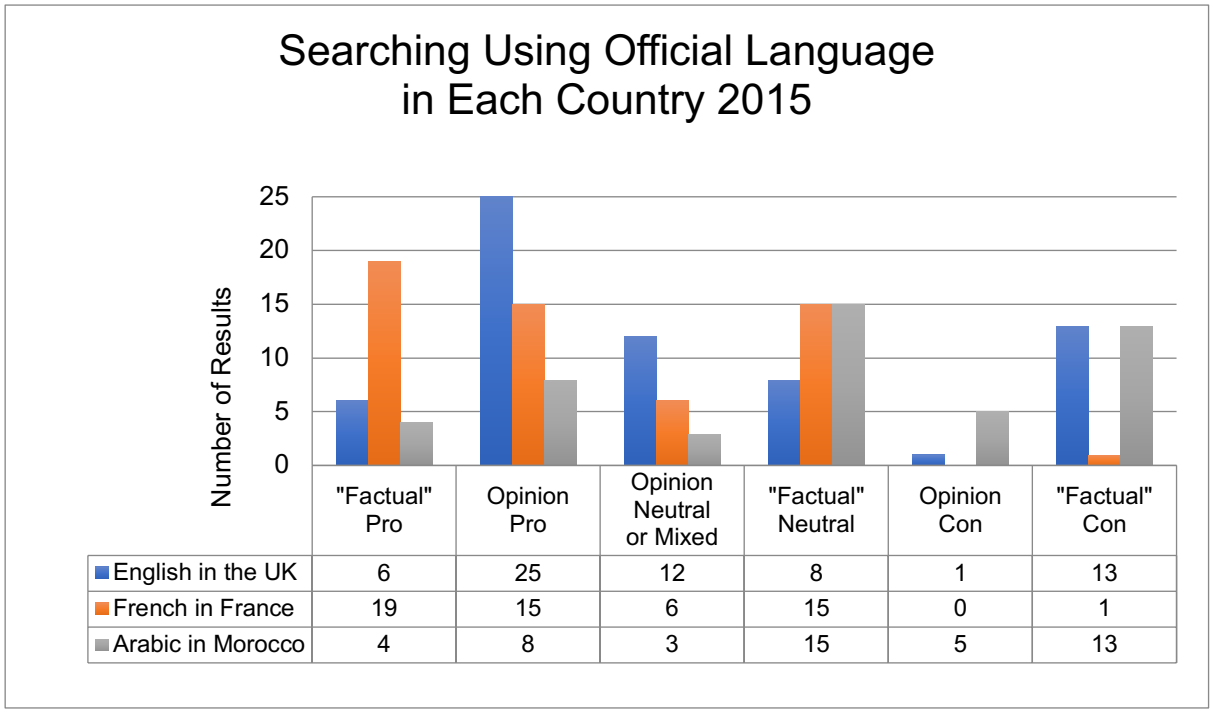
The content of the dominant Jehovah's Witnesses page is presented in a "Factual" mode, citing bible verse, and takes a Negative stance on homosexuality. Regardless of the personal beliefs of users, it is noteworthy that searching in a range

of languages and countries in 2017 produced, in this instance, a relatively monocultural perspective to a multifaceted question. Such an example brings to the fore the above statements of Schmidt and Cohen, and in particular their narrative that “People who try to perpetuate myths about religion, culture, ethnicity or anything else will struggle to keep their narratives afloat amid a sea of newly informed listeners. With more data, everyone gains a better frame of reference” (35). Irrespective of the particular content provided by search engines, one of their fundamental characteristics is their ability to return a number of different results for each single query. That one of the contrasts between the 2015 and 2017 results is a lack of overall diversity and a dominance of a single narrative is troubling. The standardisation around a single “Factual” discourse fits with Schmidt and Cohen’s narrative, but due to the automated nature of search engines, the particular message that dominates reflects the structure of the web, rather than the validity of its claims.

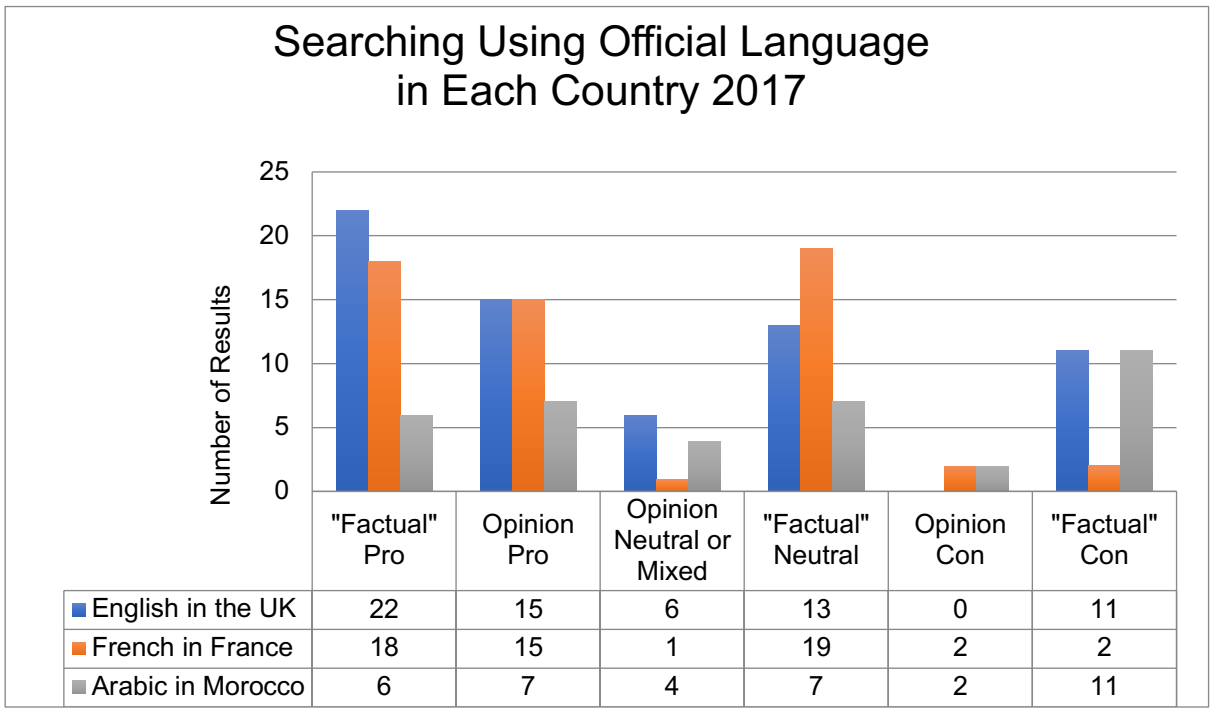
6.1.1 Search Using the Official Language in Each Domain 24 Months Later

In comparing results for each official language in each domain there were some clear changes. The number of results rated as “Opinion Neutral or Mixed”, sites such as Debate.org (see fig. 8) that frame various Opinions, was much lower for English in the UK and French in France than in 2015 (see fig. 36 and 37). One of the largest shifts can be seen in how the discourse of English results changed, from Opinion-based results in 2015 to “Factual” ones in 2017. The 2017 English results were also more Positive about homosexuality than in 2015 (2015: Pro: 31, Con: 14; 2017: Pro: 37, Con: 11). However, this does not take into account the *order* of the ranking; as shown in fig. 38, the two top results for [is being homosexual wrong?] searched in

English in the UK 2017 were two pages from Jehovah's Witnesses website discussed above. These two pages are written in a "Factual" mode and condemn homosexual acts; therefore, the graphs of fig. 37 might provide an indication of the overall landscape, or episteme, of a search context but given that the top results go *against* the general tone, conclusions are difficult to draw out.



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
English in the UK	6	25	12	8	1	13
French in France	19	15	6	15	0	1
Arabic in Morocco	4	8	3	15	5	13



	"Factual" Pro	Opinion Pro	Opinion Neutral or Mixed	"Factual" Neutral	Opinion Con	"Factual" Con
English in the UK	22	15	6	13	0	11
French in France	18	15	1	19	2	2
Arabic in Morocco	6	7	4	7	2	11

Figures 36 and 37. Graphical representations of the results from using languages in the country in which it is the official language, in 2015 and 2017, respectively.

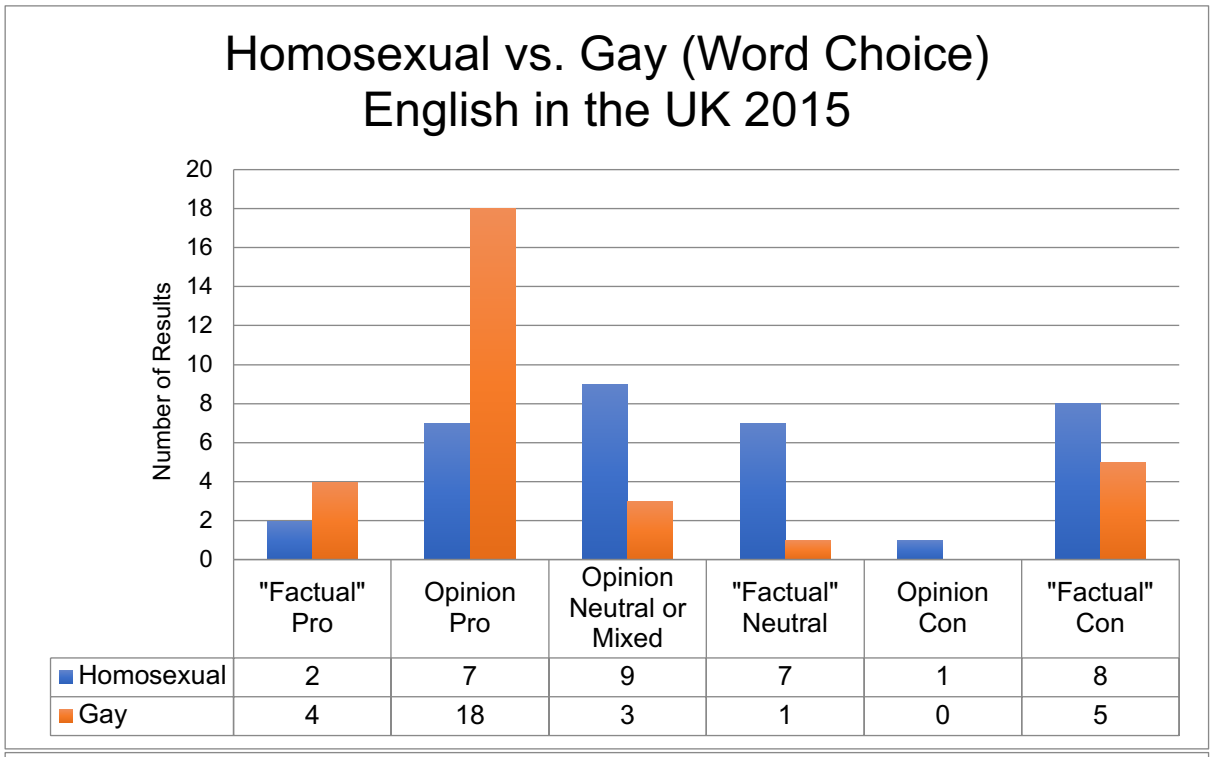
Figure 38. Search using English in the UK 2017 for the query [Is being homosexual wrong?]. Despite "Factual" Pro having the highest number of results, when all first-page results were taken into account, the top two ranked results were "Factual" con. Accessed 22/11/2017.

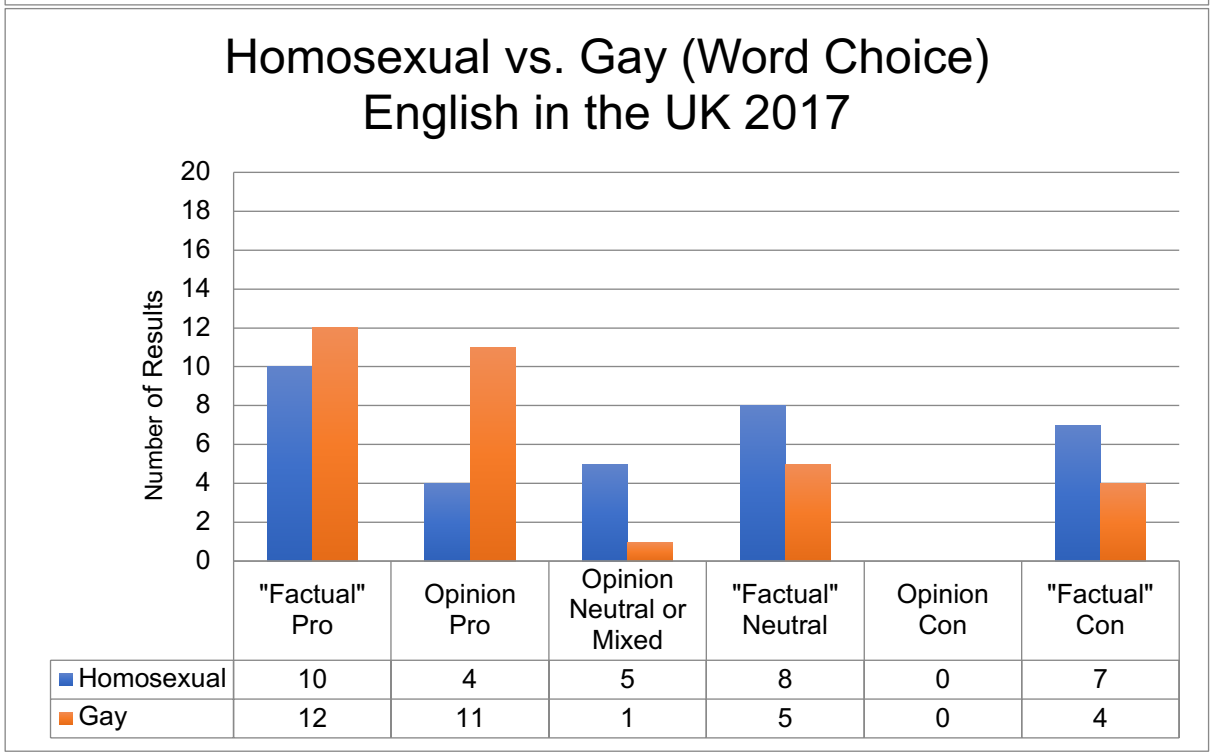
The next two subsections address changes in word choice and phrasing. In order to reduce the number of variables being compared, such as a change in the way particular translations were interpreted, these two sections focus exclusively on the context of searching using English in the UK.

6.2 Word Choice 24 Months Later

For this context, word choice variation, using the word [gay] or [homosexual], had generally standardised (see figs. 39 and 40). In 2017, there was still a significant difference between the number of Opinion Pro results that were returned when switching between the words [gay] and [homosexual]. However, queries using the word [gay] still resulted in a high number of same-sex Positive results but these were

split between “Factual” and Opinion modes of discourse. There were still differences between the kinds of results returned depending on word choice but these differences seem to have become more minimal.

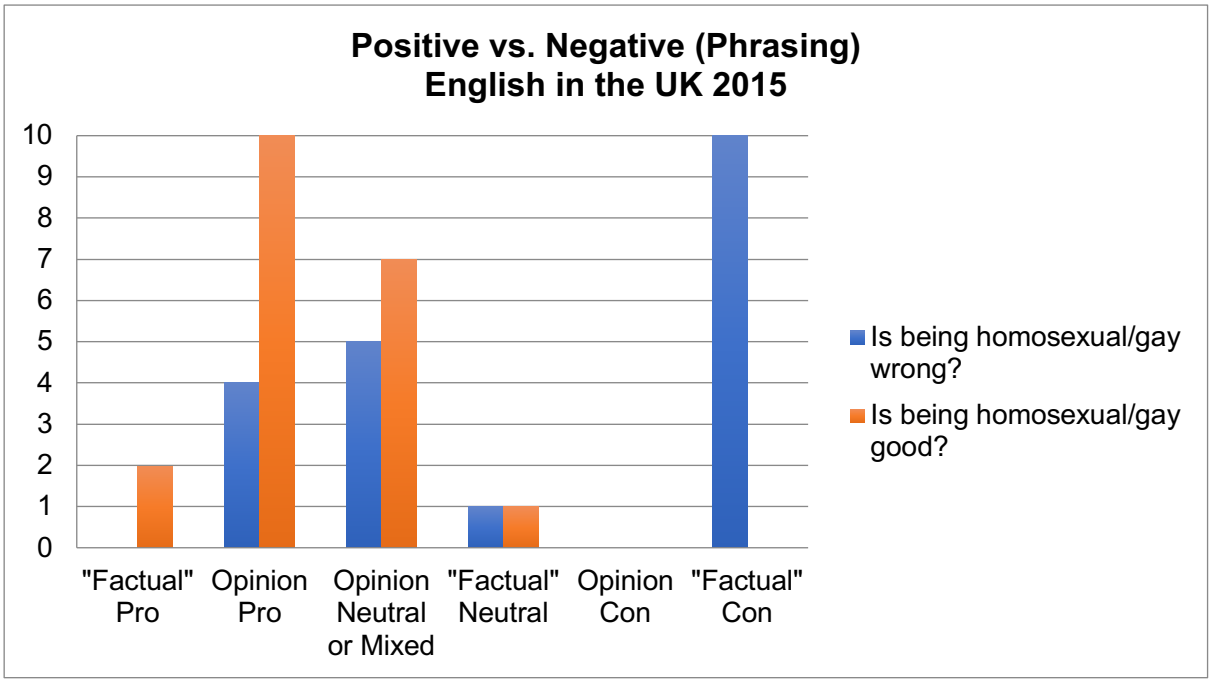


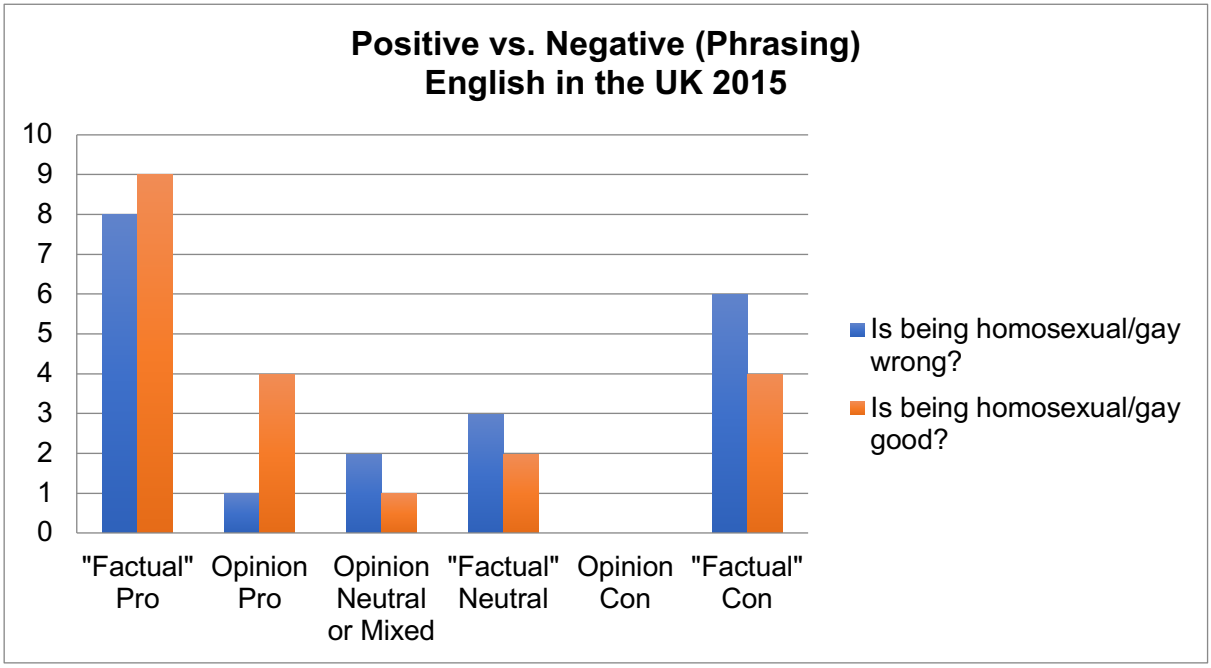


Figures 39 and 40. Graphical representations of the different results for word choice within a query, using either [homosexual] or [gay], using English in the UK, in 2015 and 2017, respectively.

6.3 Phrasing 24 Months Later

Figs. 41 and 42 clearly show this trend of standisation. The results in 2015 showed, dramatically, that the way in which a query was phrased could have a significant impact on the webpage results presented. Not only were results for [homosexual] Negative but they were presented as “Factual”, while queries using the word [gay] returned Positive pages framed as Opinion. The most probable cause was that the results containing words such as “good” tended to be Positive and casually written, while those that used the word “wrong” were citations from a more formal and Negative discourse. Although in 2015 Google’s algorithms were much more complex than the initial examples in chapter one regarding AltaVista’s methods of searching for exact word matches, this logic seemed to remain a dominant part of the ranking. The difference between the highly polarized 2015 results and the same searches in 2017 is significant. The way of phrasing a query in 2017 had far less effect on the kinds of pages produced; the same kinds of alignments as were found 2015 were present but to a less dramatic degree.





Figures 41 and 42. A graphical representation of the phrasing of queries, for example [Is being homosexual wrong?] or [Is being homosexual good?], using English in the UK, in 2015 and 2017, respectively.

It seems that whereas the results from 2015 embodied a more personalised relevance (Group B) that deployed polarized results for different kinds of searchers,

the 2017 results represented what might be considered as a move towards an editorial attempt to objectivity (Group A).

The difficulty with studying this kind of algorithmic behaviour is that robust generalisations cannot be drawn from these particular differences. Even if further evidence were available it would not help establish lines of causation. Are these results an indication of a wider acceptance of homosexuality, drawn from user behaviour, or evidence of Google asserting an editorial hand into their algorithmic modifications? These kinds of questions cannot be properly answered; in addition, even if researchers could be sure that, across a range of topics, differences in context were having less of an impact, it would be wrong to conclude that this would be a unidirectional change. It might be that a repeat of the same tests in another 24 months would indicate that contextual factors are, once again, creating a heterogeneous search environment. The nature of studying Google's algorithmic judgements means that the object of analysis is always changing and the lines of causation are never clear. It would also be incorrect to assume that Google's engineers would always be able to explain the ranking of particular search results, given their production by highly complex algorithmic systems. Changes in Google's results might represent changes to their ranking methods, changes to the state of the web, the changing behaviours of web users, or a combination of all three factors.

However, something that is known is Google's increasing reliance on machine learning neural networks, such as RankBrain. As already stated, RankBrain's original function was to use machine learning to provide effective predictions in unknown contexts; it may well be then, that RankBrain is causing the standardisation across different contexts by providing the results from more common search contexts. This would lead to a standardisation of results over a set of different contexts, which is

what the 2017 results showed. The importance of such a change would be significant considering, as outlined earlier, machine learning is far more difficult for even the engineers themselves to understand. RankBrain's machine learning may lead down a path that enforces Schmidt and Cohen's culturally imperialist sentiments, regarding Malawian witch doctors and Yemeni tribal elders; it is also a path that allows them to avoid taking responsibility for the kinds of results that become dominant. However, what should be clear from some of the examples highlighted above is that, due to the open-ended nature of query topics, search results are not simply a matter of truths and falsehoods; the discursive modes in which arguments are presented, as either "Factual" or Opinion-based, may have as much of an impact upon the individuals who turn to search engines to provide them with answers. Machine learning extends the current paradigm in which technology companies shirk responsibility. Such techniques may well lead down a path in which the global informational landscape is generated in a way that makes evidence-based critique impossible. At present, algorithmic decisions are hard to critique but they can be analysed in a way that takes into account the attitudes and aims of a company's engineers; machine learning does not afford the same insight and we should be cautious of a direction that might lead complex cultural issues to be framed without human intervention or responsibility. The destination of such a direction is a situation in which, upon being asked *why these results?* engineers might only be able to shrug.

Conclusion

This chapter began with a discussion of the unavoidable methodological difficulties that studying search results brings. The nature of search engine results presents a challenge even to the simplest acts of observation, let alone aggregation, comparison, analysis and the process of making evidence-based conclusions. That search engines rely on algorithmic criteria, the weighting of which is hidden to a user, presents a range of problems to various disciplinary approaches. It is important that academic discourse establishes robust methods of critique, for the computational logic that underpins search engines is becoming widespread throughout many areas of digital culture, including financial markets, news aggregation, and the social media feeds of a large number of individuals. In addition, the increasing reliance on neural networks by Google, and other globally significant technology companies, increases the difficulties of academic study, as well as raising questions regarding responsibility.

Associated with these methodological debates is the wider question of evaluation, at the heart of which lies the question, *what should search engine results be?* Describing theorists as belonging to either Group A, which emphasises the democratic or objective nature of search engines, or Group B, those who treat personalisation and contextual specificity as the fundamental logic of search engines, highlights the complex and polarised discourse underpinning important public debate. The key difficulty in reconciling the positions of Group A and Group B thinkers is that their attitudes refer to both the ideals that search engines should strive towards and the aspects they consider to be the *fundamental nature* of the technosocial amalgam that search engines represent. Outlining these positions

shows how subjective cultural attitudes are deeply embedded in the ways of describing technical function; this, in turn, impacts the kind of evaluative judgments that can be made by academics, programmers, legal institutions such as the EU's Antitrust Commission, and individuals in the public sphere.

The tests carried out in 2015 and 2017 demonstrate that search results differ depending on a range of contextual factors or, to use Google's terminology, signals. Due to the algorithmic nature of search engines, the results should not be taken to justify any generalising claims regarding the kinds of results provided in various countries and for different language users. That the results *are* contextually dependent and that the range of results can vary considerably is a significant enough conclusion. Demonstrating that the results tested changed between 2015 and 2017 should reinforce a hesitancy to claim absolute knowledge regarding the kinds of content provided by Google. In addition, due to their dynamic nature, there is no way to visit past states of search engines in the way that the Internet Archive's Way Back Machine allows users to visit previous snapshots of static sites. This presents a significant challenge to academic research; if researchers do not capture a piece of information at the time, there is no way to revisit previous results.

The prominence of the Jehovah's Witnesses webpage in 2017 provides a clear example of an unexpected outcome for which the original methodology of 2015 was not designed. In 2015, there were no examples of webpages that dominated across different languages; the top three results varied dramatically across each context and even Wikipedia pages from different domains did not rank as consistently highly as they did in 2017. For this reason, the original experimental design focused on the results of Google's first page, as a whole, and the specific top-ranking pages were not tracked to see if they featured in other search results. It

would have benefited the later comparison if the repeated instances of particular pages had been recorded, as this would have allowed the number of times they appeared as a result in different contexts to be analysed. However, there is no way to return to this earlier state to refocus the experiment. Such a conclusion naturally calls for research methodologies of the future to be co-created by a network of interdisciplinary researchers in order to capture a range of information and highlight various areas. However, as previously noted, researchers in the social sciences have been wary of studying these kinds of instances, where the procedures are algorithmic and the underlying data is proprietary and thus unavailable to researchers. The relationship between technology companies and academic institutions was briefly touched on in chapter three, regarding the AOL query log data, the release of which caused a number of damaging repercussions, of a social, economic and legal nature. An automated way of scraping search result data may well be useful but at present is not feasible given that Google reserves the right to reject the use of certain Application Programming Interfaces (APIs). During both the 2015 and 2017 experiments, searches were frequently stopped by Google due to “unusual traffic” requiring the completion of a Captcha test (see fig. 43) even when using a VPN throughout.

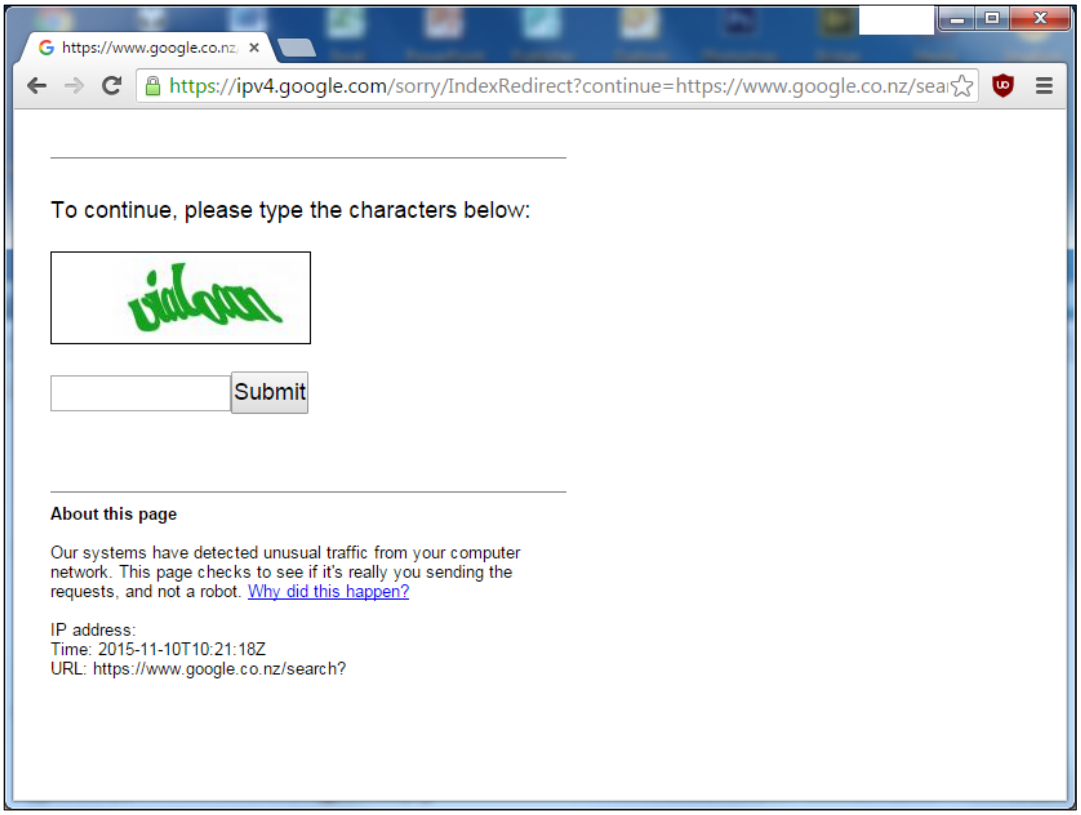


Figure 43. An example of a Captcha test enforced by Google in response to “unusual traffic”. Accessed 22/11/2017.

Finally, the difficulty for the average user to see the confines of their online context has recently increased. In an official Google blog post published 27 October 2017, Evelyn Kao, product manager for Search, announced a change in the way users could control their geographical domain location. Although this study relied on a VPN to control the location that was used by Google to determine results, before 27 October 2017 users could easily alter their geographical domain simply by manually typing a different country code top-level domain (ccTLD), at the end of their URL. For example, if a user in Germany wished to see what their results would be if searched in France, they could navigate to www.google.fr, or the UK by visiting www.google.co.uk, or Morocco by entering www.google.co.ma. This provided a simple method for individuals to perceive how different results were served

depending on the contextual signals sent to Google. After the change on 27 October 2017, typing any ccTLD for www.google.“…” is ignored and the user is redirected to the domain that matches their geographical signal. For example, without using a VPN, trying to navigate to www.google.fr, when based in the UK leads Google to force the user to www.google.co.uk, even when using the URL command No Country Redirection, by appending /ncr to the end of the URL. The option to change domain is available in “Search Settings” however, even when changed the results *still* draw from the location of the user’s IP address, rather than the chosen domain. This might seem like a niche issue that only concerns researchers, however, I believe that if web results are to be localised and personalised then users should have the right to see how their results are different to another user, in another context. The above discussion regarding *unimaginable communities* was written before Google stopped users from changing their location by changing their ccTLD and represented a metaphorical description regarding how ways of thinking and acting online are always contextual. However, since 27 October 2017, Google’s decision to give users even less control over their online context highlights its practical implications for the future of the web. The direction of Google’s search strategy, as it appears at the end of 2017, seems not only to favour the global segregation of search results but also a direct attempt to prevent users from seeing the digital barriers that are enforced upon them.

Chapter Five:

Google and Advertising: Digital Capitalism in the Context of Post-Fordism, the Reification of Language, and the Rise of Fake News

Introduction

Google's dominance over the web allows it to dictate various norms and practices that regulate the state of contemporary capitalism online. The way in which Google operates as a company and generates revenue is often sidelined in academic discussions regarding the cultural implications of how its search engine functions. Almost 90% (Alphabet 22) of Google's revenue is derived from advertising, despite Larry Page and Sergey Brin's original 1998 academic paper regarding Google in which they argue that advertising produces mixed motives that make it an unfeasible way to fund search engines.

This chapter outlines how Google's model of advertising reflects and encourages wider changes in capitalism as it shifts from its twentieth-century Fordist incarnation to contemporary Post-Fordist arrangements of labour. In doing so, this chapter analyses Google's two main advertising systems, AdWords and AdSense, and proposes that these financial models have significant effects upon online discourse. A discussion of AdWords details some of the tensions between the local

and the global that develop when tracing flows of information and capital, specifically highlighting Google's impact on the decline of online language diversity. An outline of AdSense demonstrates how Google's hegemonic control prescribes which parts of the web can be monetised and which remain unprofitable. In particular, in drawing from existing studies, this chapter provides evidence that Google's AdSense programme, along with Google's relationship with Facebook, incentivised the rise of fake news in the 2016 US presidential election.

As has been highlighted throughout this thesis, search engines have been addressed from a wide range of academic perspectives including, but not limited to, computer science, law, politics, information retrieval, and new media studies.[1] However, the role that Google's advertising business model plays within contemporary capitalism is rarely addressed directly. This topic is crucially important because it impacts on a wide range of phenomena that might otherwise be considered non-economic, such as general online language use, and the incentives underpinning a range of content, such as fake news. In order to situate these issues within broader trends in contemporary capitalism, this chapter draws on Post-Fordist theory, which is outlined below, placing it within a digital context. This perspective highlights the extensive influence that Google's revenue model has upon digital culture. Emphasising Google's business activities recontextualises the arguments of the previous chapters in terms of their underlying economic incentives. Doing so provides a more complete picture of Google's activities and their cultural, political and economic effects.

[1] Examples that exemplify such disciplinary approaches are as follows: for computer science, see *Google's PageRank and Beyond: The Science of Search Engine Rankings (*Langville and Meyer); for law, see *The Googlization of Everything (And Why We Should Worry)* (Vaidhyanathan); for politics, see *The Dark Side of Google* (Ippolita) and *When Google Met WikiLeaks* (Assange); for information retrieval, see *Web Search Engine Research* (Lewandowski ed.); for new media studies, see *Google and the Culture of Search* (Hillis et al.) and *Society of the Query Reader: Reflections on Web Search* (König and Rasch eds.).

1.0 The Economics of Google

Alphabet, Google's holding company created in 2015, is one of the most valuable companies in the world.[2] It has a market value of over $500 billion and in 2016 generated a revenue of $77 billion. Ask someone what Google does and they will likely reply that it is a search engine company. However, a more accurate description is that Google is an advertising company. 88.7% of Google's revenue comes from advertising (Alphabet 22); although, as will be outlined below, Google's modes of advertising deviate significantly from any existing forms of traditional advertising. The economic success of such a shift is producing dramatically widespread effects within many areas of society. This chapter addresses two in particular: the reification of online language and the rise of fake news. There are many other important impacts of Google's advertising programs, however, focusing on these two issues demonstrates the broad scope on which such a narrow economic model operates.

Google has two main advertising ventures. The first of which is "Google properties", the service for hosting advertisements built into its own products (its search engine and Gmail, for example) the most significant part of which is AdWords. The second is "Google Network Members' properties", a brokerage service that runs advertisements on third-party websites, the most significant part of which is AdSense.[3] AdWords and AdSense will be outlined separately, as they each have different impacts, and one pertinent consequence of each will be highlighted, the reification of language and the rise of fake news, respectively. 71.3% of Google's

[2] Alphabet has been listed as the world's most valuable company on a number of occasions; see "Google Just Passed Apple As The World's Most Valuable Company" and "Google Just Passed Apple As The World's Most Valuable Company (Again)" (Solomon).

[3] For clarity, this chapter relies on synecdoche by using "AdWords" to refer to "Google properties" and using "AdSense" to refer to "Google Network Members' properties". There are other smaller properties detailed in Alphabet pages 23 and 24, respectively. However, these smaller properties follow the systems developed by AdWords and AdSense and so, in the main, can be ignored.

revenues comes from advertising on Google's own sites and is mainly derived from AdWords (23). AdWords is an auction process that Google operates to allocate paid results (referred to by Google as sponsored results) to search engine queries, which sit separately on top or to the side of unpaid results (referred to by Google as organic results, see fig. 1).

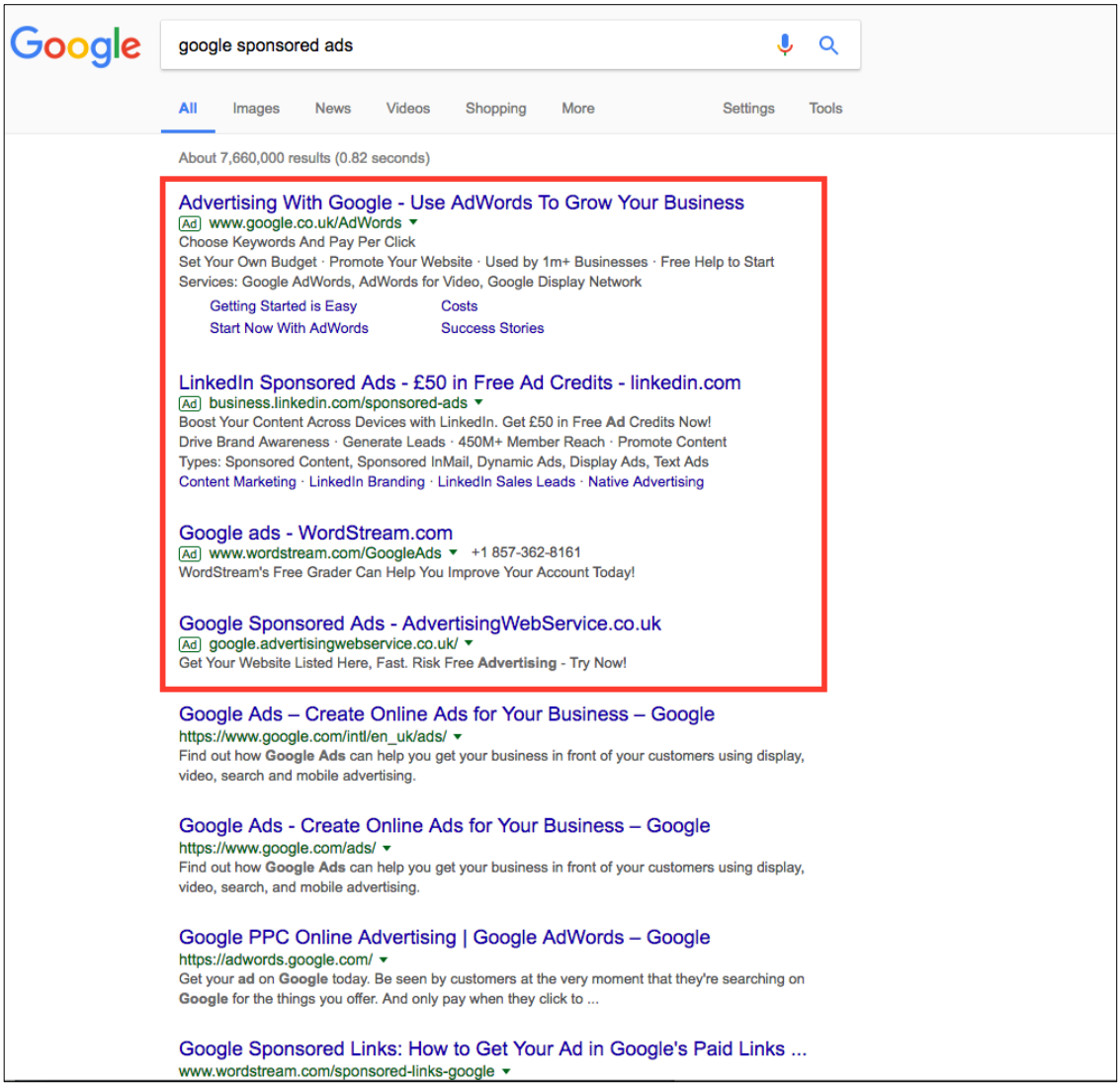


Figure 1. Google's sponsored results. The results in the red box are the sponsored results; the results outside of the red box are the "organic" results organised by the PageRank algorithm and other factors. Drawn by the author.

These sponsored AdWords results are visually delineated from the organic results produced by the PageRank algorithm and other factors (see fig. 2 for a visual history of how these sponsored results have been displayed).

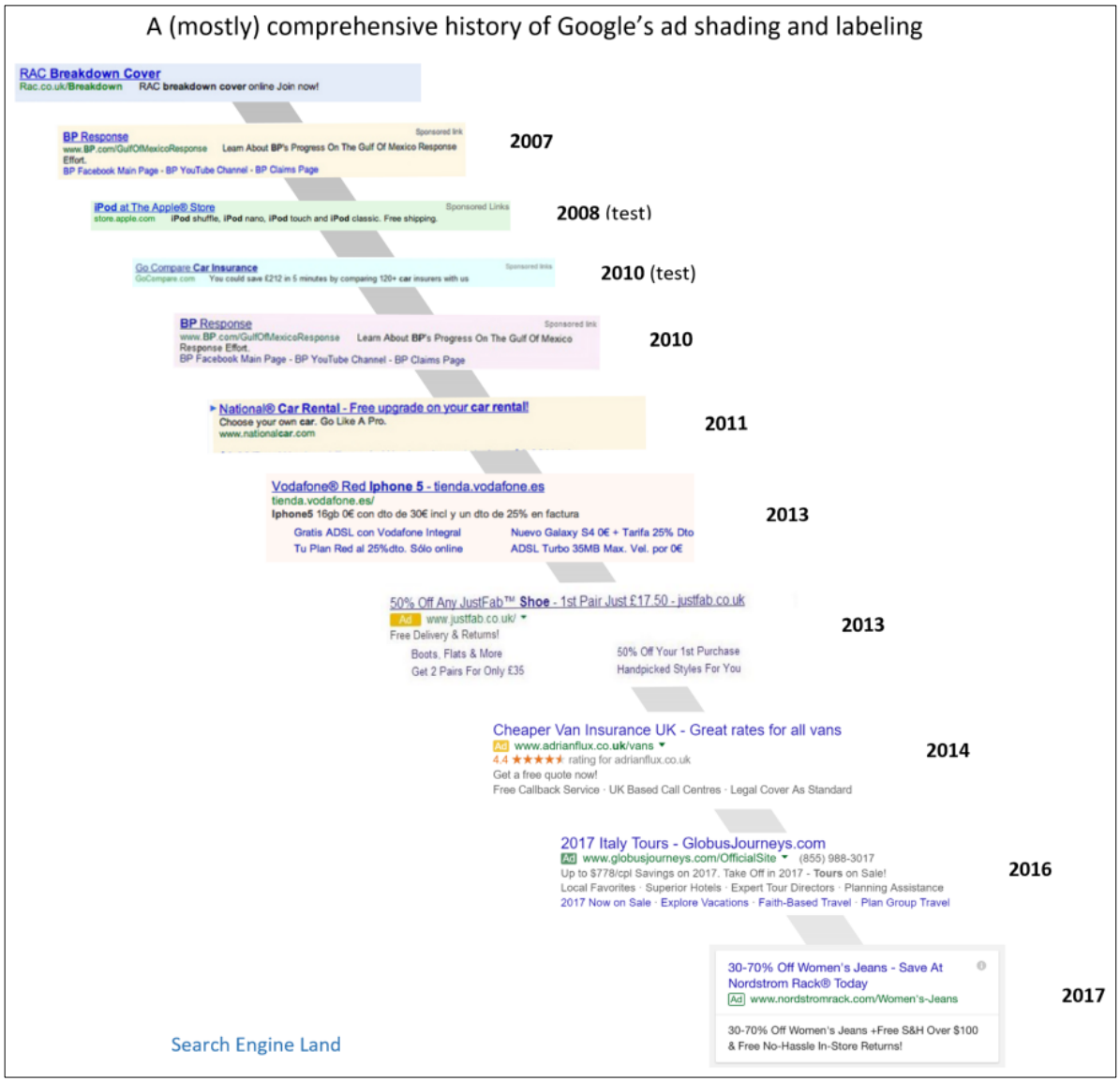


Figure 2. "A (mostly) comprehensive history of Google's Ad shading and labeling" taken from Marvin.

In addition to this program, 17.4% of Google's overall revenue is derived from non-Google sites on which Google hosts third-party advertising content using its AdSense programme (24). AdSense is Google's method of linking third-party

advertisements to relevant third-party content, such as blogs or news sites, and displaying advertisements alongside selected content in digital billboards. Although there are other companies that provide advertising for third parties, Google AdSense is by far the largest. It was reported in *The New York Times* that “In the first quarter of 2016, 85 cents of every new dollar spent in online advertising will go to Google or Facebook” (Herrman). The second section of this chapter outlines how these two companies reinforce their joint dominance.

The remaining 11.3% of Alphabet’s revenue represents the sales of apps and media content in the Google Play store as well as other smaller ventures such as certain Google branded hardware, for example sales of Google Chromebooks and Pixel smartphones. None of Alphabet’s other activities “meet the quantitative thresholds to qualify as reportable segments; therefore, the operating segments are combined and disclosed […] as Other Bets” (Alphabet 21) These smaller subsidiaries, known as other bets “Access, Calico, CapitalG, Nest, Verily, Waymo, and X,” and other initiatives (21) have a combined revenue of $809 million, which is only 0.9% of Alphabet’s total revenue, and have combined operating losses of over $3,578 million (74). These divisions focus on a range of projects from Calico and Verily’s biomedical research into extending the human lifespan, Waymo’s self-driving cars, and X’s Google Glass augmented reality headset. Whilst these kinds of projects are covered more frequently in the popular press they do not contribute to Google’s financial success.

In summary: Google generates almost all of Alphabet’s revenue; almost all of Google’s revenue is made from advertising; the majority of this advertising revenue comes from AdWords, i.e. sponsored links included in search engine results; finally, Google spends a great deal of its revenue on smaller ambitious ventures. This

chapter draws from Post-Fordist theory to provide a historical and theoretical context for Google's place in contemporary digital capitalism. In doing so, this chapter demonstrates the ways in which the Internet and the web have played a major role in the changing flows of information, labour, and capital and outlines how at the heart of this change is Google: both the company and its products. The first section of this chapter focuses on AdWords, arguing that it constitutes a global linguistic market and typifies a number characteristics of Post-Fordist capitalism. This section also addresses how Google's financial model contributes to the decline of language diversity online by incentivising the use of more profitable languages over others. The second section of this chapter focuses on the ways in which AdSense shapes online discourse and dictates particular norms. In particular, this section draws out the reciprocal links between AdSense and Facebook and demonstrates how Google's mode of advertising facilitated the rise of fake news during the 2016 US presidential election. Before addressing these two topics, the following section establishes the framework of Post-Fordist theory, in order to contextualise Google's activities within the broader shifts of contemporary capitalism.

2.0 The Context of Post-Fordism

Google's modes of advertising represent a sea-change from traditional twentieth-century advertising which, in turn, operates within a much larger and more general shift away from twentieth-century modes of capitalism. This chapter draws from a particular group of thinkers, a group of Italian Neo-Marxists, loosely connected to the "workerism" (*operaismo*) movement during the 1960s and 1970s, in order to better contextualise contemporary digital capitalism. The work of these thinkers – Paolo Virno, Michael Hardt, Antonio Negri, and Maurizio Lazzarato are a few notable examples – has come to be known as Post-Fordist theory.[4] This designation of Post-Fordist theory represents a number of different, although related, attitudes towards the way in which capitalism has changed since the second half of the twentieth-century, specifically in regard to the role of work. Post-Fordist theory traces the decline of the dominant kind of capitalism in the early twentieth-century, Fordism,[5] while theorising and analysing the post-industrial modes of capitalism that they describe as Post-Fordist. These models of Post-Fordist labour relations, which stress the importance of cognitive, flexible, and precarious labour, are key to understanding Google's influence on contemporary capitalism in a digital context. Various other thinkers outside of this Italian Neo-Marxist group have described the current mode of capitalism emphasising similar characteristics under a range of names: "Empire" (Hardt and Negri), "Late Capitalism" (Jameson), "PostCapitalism" (Mason),

[4] The collection *Radical Thought in Italy: A Potential Politics* (Virno and Hardt ed.) goes some way to providing a comprehensive overview of the important Post-Fordist theorists, although as is noted in its introduction three key members, Franco Berardi, Sergio Bologna, and Giuseppe Cocco, were not included for various reasons.

[5] Fordism emphasises standardisation, de-skilling of the workforce through assembly-line manufacture, and the linking of wages to prices of products in order to ensure that workers could function as consumers of their products. For a more detailed definition and relationship to later forms of capitalism see *Post Fordism: A Reader* (Amin).

“Capitalist Realism” (Fisher), “semiocapitalism” (Berardi), “Cognitive Capitalism” (Boutang). Although these related approaches use different terminology, the various characteristics highlighted complement a description of contemporary capitalism as Post-Fordist.

The effects of Post-Fordism are numerous and many directly relate to Google’s role in contemporary capitalism. To attempt to address all of these characteristic changes goes beyond the scope of this thesis. However, it should be noted that Google’s advertising ventures of AdWords and AdSense operate in concert with a range of other contemporary issues. These include: the increasingly precarious nature of employment; the dissolution of clear boundaries between work and free time, as well as between paid and unpaid work; the diminishing solidarity, rights, and freedoms of workers; the increasing time spent working, in each working day as well as an increasing age of retirement; the homogenisation of different types of work through the use of information technologies; employing automation to replace workers; the changing nature of digital commodities that turn many product based industries into service ones. The list is only indicative, rather than exhaustive, aiming to provide a sense of Post-Fordism’s extensive nature. An exaggerated example of a day that exemplifies Post-Fordist labour relations might be as follows: a woman catches an Uber to her timeshared office where she works as a digital brand consultant. Her work consists of managing Twitter likes and increasing Facebook engagement. She orders her lunch via an app, which is delivered by a part-time student working as a Deliveroo rider. She spends her evening watching Netflix, a subscription television service, as she does not have room to keep DVDs in her small Airbnb. That evening she spends an hour talking to a Chinese student over a language learning service, like italki, not to make money but in exchange for credits

to be redeemed at a later date, once she finds the time to start learning Spanish. Not all work is like this, far from it; material labour is still a major part of contemporary work around the world. In addition, this example only highlights the visible changes that might seem to only affect a niche group of people. However, this chapter aims to show that Google's business draws *everyone* online into various immaterial labour arrangements with far reaching consequences, many of which are difficult to detect. There are numerous dimensions to such an arrangement and different people are implicated into Post-Fordist labour relations to different degrees. Focusing on how AdWords reifies language online and how AdSense incentivises fake news demonstrates two examples in which all web users are impacted by Post-Fordist effects, even if their lives could not seem further from that of the example outlined above.

3.0 AdWords: Organic vs. Sponsored Results

> At least in terms of revenue generation, Google's core business isn't facilitating searches, it's selling advertising space — or rather, selling our attention to advertisers and managing both the price it charges for access to our attention and the relative visibility of those advertisements. (Vaidhyanathan 26)

Vaidhyanathan's above quotation is useful because it draws our attention to Google's profitability as a company. That the majority of Google's revenue comes from the advertising through its search engine should be surprising, given that the original plan for Google's search engine was diametrically opposed to advertising. Despite the plans of the founders for Google to remain in the academic realm, Google generates revenue when users click on advertisements, not when users find successful answers to their queries. As Steven Levy describes, "In their original academic paper about Google, [Larry] Page and [Sergey] Brin had devoted an appendix to the evils of conventional advertising" (*In the Plex* 84). Their academic paper argued that their method, using the PageRank algorithm, was far more accurate than existing search engines that relied on advertising *specifically because* it did not bias results in order to make a profit. Their approach required that their search engine be "transparent and in the academic realm" because, as the founders explain:

> advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers. […]

> Furthermore, advertising income often provides an incentive to provide poor quality search results. For example, we noticed a major search engine would not return a large airline's homepage when the airline's name was given as a query. It so happened that the airline had placed an expensive ad, linked to the query that was its name. A better search engine would not have required this ad, and possibly resulted in the loss of the revenue from the airline to the search engine. In general, it could be argued from the consumer point of view that the better the search engine is, the fewer advertisements will be needed for the consumer to find what they want. This, of course, erodes the advertising supported business model of the existing search engines. (Brin and Page)

However, early on in their business, Google started using advertisements to fund their search engine. These advertisements, sponsored links, have always been kept separate from the organic links. In their paper, Brin and Page specifically took aim at search engines that mixed their results together so that users could not see which of the results had been paid for and which were freely chosen by the search engine. In this regard, Google have not gone back on their original statement; however, the problem remains: if a set of results is good enough, a user will never need to click on the sponsored link. In the original vision outlined by Brin and Page, advertisements are always an indication of failure, but today represent the overwhelming majority of Google's revenue. The following section outlines how this perspective of success vs. failure is deceptive and provides a different lens with which to describe the complex relationship between Google's search engine and advertising.

3.1 AdWords: The First Global, Real-time, and Multilingual Linguistic Market

This section provides a specific outline of how AdWords operates in order to demonstrate the close links between Google's search engine and advertising. In particular, this section draws on the work of Frederic Kaplan who argues that AdWords constitutes a form of "linguistic capitalism" (58), in the tradition of Post-Fordism; the conclusion of which is that Google's mode of advertising is having a widespread effect on all language usage on the web. Even if users are not explicitly altering the language they use online, anyone who uses the web communicates in a context where economic value alters every part of their linguistic landscape. Users may be completely unaware of this process of linguistic reification but still navigate an uneven digital space in which there are economic incentives that prioritise some words and ideas and deprioritise others. In addition, the economic value of different languages is not the same, and therefore, as discussed later, each language group is affected to a different extent.

AdWords is the auction system that provides advertising in the form of sponsored results that fill the top or side of a Google search result. An auction occurs every time a query is searched and balances the amount of money automatically bid by a company, against an automated quality score, given by Google. If low-quality scores are given or if there has not been an advertisement placed that is deemed relevant to the query, Google will not provide an advertisement. If a user clicks on one of these sponsored advertisements, the company being advertised pays Google; if not, no money is exchanged. Therefore, both Google and its business customers have strong economic incentives for the

advertisements to succeed. A key aspect of AdWords is designing advertisements with specific kinds of search queries in mind. Kaplan explains that, “First, advertisers select a keyword—for instance ‘vacation’—and define the maximum price they would be ready to pay if a user arrives on their site by clicking on the link of the ad” (58). Keywords selected by advertisers are then used interchangeably with other similar words selected by Google. The advertiser must also select the “product or service [they] wish to advertise” from a predetermined list, the language they wish to advertise in, and the geographical locations they wish to target with their advertising. Google have guidelines for prohibited AdWords content,[6] which mostly relate to more general country-specific laws.[7] In addition, the process of having to choose from a pre-established list of products and services means that many taboo grey areas are condensed into more general subjects and thus implicitly censored. After this process is completed

> Google associates a quality score with the ad. This figure, ranging from 1 to 10, evaluates the global “quality” of the ad, which is computed through a complex combination of various factors, including the relevance of the text ad regarding the keyword, the average number of clicks on the ad, and the performance and quality of the linked website. This score measures how well the ad is working. (58)

This measure of “quality” takes into account a judgement of the advertisement (clarity of expression), its relevance to the website it links to (an advertisement for swimming goggles should lead to a sports equipment shop rather than a public

[6] See Google “AdWords policies”.
[7] For example, it is illegal to advertise online gambling websites in the US but legal in the UK and Google follows these geographical distinctions online. See Google “Gambling and Games”.

swimming pool), and the quality of the destination website (based on Google's usual metrics including the layout of keywords and the link score generated from other websites linking to that site). Whereas traditional advertising might aim to change someone's mind or introduce them to a new idea, AdWords advertisements are an attempt to reflect the existing perspective of an individual. AdWords advertisements, in aiming to be *relevant*, need to replicate the current outlook of a user, or successfully predict their context, in order to be given a high-quality score. Finally, the "rank of an ad is calculated by multiplying the bid times the quality score" therefore an "ad with a good score and medium bid can overcome a less efficient ad with a higher bid. Eventually, the price paid by the advertisers is not their maximum auction offer but a slightly lower price, one computed on a second-price auction model" (59). This financial model means that small advertisers can compete with larger ones if they can offer a higher quality advert, as judged by Google. So, for example, in a search for "craft ale" (see fig. 3), Amazon might have set the highest bid for that phrase but it is listed underneath two smaller but more specialised craft beer sellers, with higher quality scores. This approach prevents companies with deeper pockets outbidding smaller but more relevant competitors.

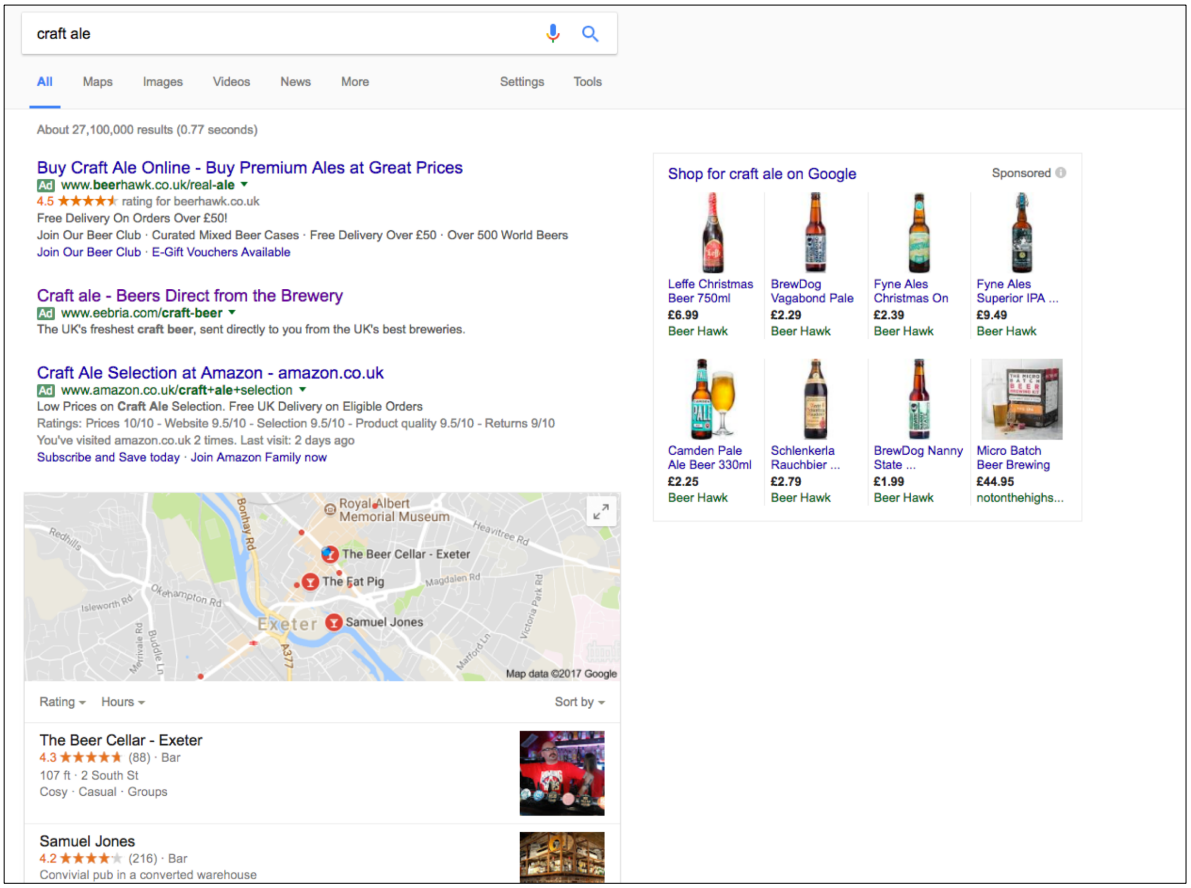


Figure 3. AdWord results for “craft ale”. The advertisement order lists the specialist and local sponsored results before Amazon’s advertisement. Screenshot by the author. Accessed 10/09/2017.

An AdWords auction occurs every single time a user searches a query using Google’s search engine. An estimate, based on Google’s statement to search engine expert Danny Sullivan, that they handle at least “2 trillion searches per year” (“Google Now Handles”), means that at the lowest estimate, Google runs a staggering 63,000 linguistic auctions per second. Kaplan describes this process as:

> the first global, real-time, and multilingual linguistic market. As a consequence, the fluctuation of the price of keywords indirectly reflects global linguistic movements. The value of some keywords like “snowboarding” or

> “bikini” varies seasonally. The increase and decrease of the word “gold” is linked with the perceived state of financial crisis. Google makes a lot of money on some very competitive keywords like “flowers,” “hotels,” “vacation,” and “love.” It also organizes bids for buying the names of famous people (“Picasso,” “Freud”). Bidding strategies vary. Anything that can be named can be associated with a bid. (59)

Google provides advertisers using AdWords with a tool that organises, suggests and estimates the cost of various words and phrases. Not only do the prices reflect real world events, but also, various companies can drive up the prices of words for their competitors. Such behaviour was raised in the 2007 court case “Google, Inc. v. American Blind & Wallpaper Factory, Inc.”. John Battelle in *The Search* details how

> in early 2003, American Blinds realised that while it owned the trademark on “American Blinds,” it didn’t own the market for it on Google's AdWords service. Competitors were snatching up the company’s trademarks as AdWords terms (they did so by paying more for them, essentially), so that when customers typed “American blinds” into Google they’d get advertisements for companies like JustBlinds.com and Select Blinds. (180)

The case of American Blinds vs. Google was described by commentators as a clear-cut trademark infringement that favoured American Blinds, however in a shock to many legal commentators, after “almost four years of litigation” American Blinds finally dropped the suit in a “‘stunning victory for Google,’ wrote Eric Goldman, an

assistant professor at Santa Clara University School of Law" (qtd. in Auchard). Google have a track record of winning legal battles that establish digital norms[8] and this case changed the conventional understanding of how legal definitions of ownership translate online. Google's legal defence outlined an aggressive position concerning the relationship between language and capital: old notions of linguistic ownership do not apply online. Not only are all words and phrases available to anyone, the auction winners are not even necessarily the highest bidder. Google's quality ranking system, when coupled with their dominance in the search engine market, means that they have become the gatekeeper of language ownership online. As language ownership is calculated and awarded anew through an auction every single time a search takes place, at least 63,000 times a second, no one can really claim ownership of language: words and their relation to entities are constantly in flux under Google's watchful eyes.

Because AdWords affects all words, not just copyrighted ones, we have a situation in which all words and ideas online are becoming commodities. The advertising model of AdWords, therefore, encourages companies into an association with language whereby a company does not connect their product with a specific slogan but to an unlimited range of words, at various times in specified locations. Many scholars have linked this expansion to the concept of the "long tail", originally popularised by Chris Anderson[9], as Levy explains: "Since Google searches were often unique, with esoteric keywords, there was a possibility to sell ads for categories that otherwise never would have justified placement. On the Internet it was possible to make serious money by catering to the "long tail" of businesses that could not buy

[8] For example, the eleven-year legal battle between Google Books and The Authors Guild, in which Google's book scanning activities were ruled as legal under fair use, see Cohen.
[9] As Anderson explains it, the "theory of the Long Tail is that our culture and economy is increasingly shifting away from a focus on a relatively small number of 'hits' (mainstream products and markets) at the head of the demand curve and toward a huge number of niches in the tail".

their way into mass media" (*In the Plex* 85). Because of this phenomenon, and in conjunction with the auction aspect of AdWords that means that uncommon words are very cheap to bid on, all words in all languages can theoretically become profitable. There is an incentive to increase the prices of popular words but also to spread the reach of a campaign to niche words that might not seem in any way commodifiable but would have a low market value. Because advertisers are only charged when a user clicks on their advert there is no cost or disadvantage to placing bids on uncommon or unlikely words. This structural logic places clear incentives on advertisers to increase the scope of their chosen words and encourages a kind of linguistic land grab. Such an expansion then has an effect on all language used online, not just trademarks or particular phrases associated with companies or products. In shifting which words and phrases become discoverable through a search engine, and which are concealed, as well as how certain language becomes received in various contexts, economics comes to structure an increasing percentage of online linguistic communication. Kaplan argues that "Even if Google's autocompletion may not be explicitly biased toward more economically valuable expressions, it nevertheless tends to transform natural language into more regular, economically exploitable linguistic subsets" (60). When born-digital content, for example online news, is written with search engine visibility in mind it is, in effect, automatically tailored towards advertising; advertisers and content creators both want to strengthen their association to the kinds of words and phrases used in search engine queries. In addition, as online content is increasingly dependent on third-party advertising, a topic that is discussed in the second section of this chapter, these two activities – bidding on search terms and writing online content that is discoverable through search engines – become enmeshed and mutually standardise

the kinds of linguistic patterns on the web.

3.1.1 Google's Institutionalisation, Data-Collection, and Advertising

This standardisation across languages is also enhanced by Google's institutionalisation of the AdWords program. Not only do companies that advertise through AdWords have access to numbers of tools and analytics, but this work is often outsourced to professional AdWords companies.[10] Google runs a certification programme which provides training, study materials and holds exams for individuals to become accredited. To keep their status as an accredited AdWords professional, individuals need to pass two of Google's AdWords exams every year. In order for an advertising company to work as a "Google Partner", they need to employ at least two members of staff who are currently accredited as AdWords professionals. The AdWords accreditation has even been added as a component of many Business Masters (MBA) degrees.[11] Google also supplies funding to those institutions awarding MBAs through their "Google Online Marketing Challenge"[12] which strengthens links between universities, professors, and students with Google AdWords and in turn strengthens Google's hegemony.

Through Google's various projects, the company has an enormous collection of data, which, when combined with their methods of tracking users' behaviour on the web ensure that Google's advertising efforts are as effective as possible. As Ken Auletta describes:

[10] Google encourages this arrangement and details advice for working with a third party, see Google "Advertiser guide".
[11] See Racer Nation Information.
[12] See Google "Welcome to the 2017 Google Online Marketing Challenge".

> It was Google's ambition, Schmidt and Page and Brin liked to say, to provide an answer to the adman's legendary line "I know half of my advertising works, I just don't know which half." To help them sort through the digital clicks, Google and other new media companies relied on what are called cookies, software files that reside on a user's browser and keep track of their activities online: search questions asked, Web pages visited, time spent on each Web page, advertisements clicked on, items purchased […] Although the cookie doesn't identify the user by name or address, it does assemble data advertisers crave and couldn't get from traditional media companies. (7)

Cookies and measurements of user interactions with search results allow Google to capture latent information that is used to further personalise advertising.

Describing Google as an advertising company (rather than a search engine that also advertises) refigures their search engine simply as a way of capturing economically useful information through mutual consent. Commentators have noted how many of Alphabet's projects that might seem very separate from Google's core business of Search, serve an important role in increasing advertising or data-collection opportunities. For example, Google Glass, an augmented reality headset released in 2013, does not immediately seem linked to the business of a search engine. However, Google was awarded a number of advertising-related patents that could be used with Google Glass. One of these patents, Pay-Per-Gaze, uses eye tracking to allow "advertisers [to] be charged a fee based on whether a person looks directly at an ad in the real world, and the fee can change based on how long they interact with the ad" (Miller and Bilton). The patent also covers the measurement of pupil dilation so that "the inferred emotional state information can be provided to an

advertiser (perhaps for a premium fee) so that the advertiser can gauge the success of their advertising campaign" (qtd. in Truong, 2013). These kinds of advances add an economic perspective with which to re-evaluate Google's mission statement: "to organize the world's information and make it universally accessible and useful" ("Google: About Us").[13] Producing a patented system that monitors and records a person's gaze and pupil dilation is a way of making existing information "useful" and provides a physiological metric with which to measure relevance. If Google Glass records that a person did not look at an advertisement for long or was not excited by it, they can change or replace that advertisement for something that sustains their gaze, widens their pupils and quickens their pulse. Other such examples can be seen in patents relating to other Alphabet ventures (the Other Bets highlighted in the introduction) such as Nest and Google Home. These patents range from "Advertising Based on Environmental Conditions" (Heath) (coordinating a range of different sensors in the home and from mobile devices), to "Coupling an Electronic Skin Tattoo to a Mobile Communication Device" (Alberth) (which consists of a microphone permanently embedded in a user's throat). Such developments allow Google to capture increasing amounts of data on and offline in order to increase the opportunities to commercialise existing behaviour.

The strategy of AdWords marks a departure from traditional advertising in a number of ways. As Levy argues, the AdWords policy "reflected the different philosophy Google brought to advertising in general. Google ads were *answers*. They were solutions. 'Ideally we wanted people to have a 50 to *100* percent click rate,' says [Tim] Armstrong [Vice President of Ad Sales at Google]" (*In the Plex* 112).

[13] Various technology commentators have also noted similar data-capture uses for many of Google's acquisitions. For example, Google acquired smart thermostat and smoke detector company Nest in 2013 for $3.2 billion. John C. Havens' 2014 article "The Connected Home May Become the Collected Home" discusses how Google might use these data for personalised advertising purposes.

Aiming for a click rate above 50% means that Google hoped for users to click on the sponsored advertisement link more often than the top algorithmically-produced organic result. In doing so, Google wanted users to place their faith in the advertised links as representing more useful or relevant answers than the search results. The way in which the advertising is so embedded in the function of Google's search engine complicates an existing notion, borrowed from older media forms such as television or newspapers, that advertising revenue financially supports a medium but fundamentally stays separate to the content of that medium. This then raises the question, is AdWords really a form of advertising at all? To further interrogate this question we now turn to Raymond Williams' critical history of advertising "Advertising: The Magic System".

3.2 AdWords in the Context of "The Magic System"

Williams' essay provides a history of advertising as a specifically contextual activity. To stress the historicity of advertising, Williams begins by dismissing a dominant conception that the history of advertising can be traced back to documents such as a "three thousand year old papyrus from Thebes, offering a reward for a runaway slave" or he adds, tongue-in-cheek, "some pleasant recollections from the Stone Age" (170). Instead, advertising "was developed to sell goods, in a particular kind of economy" (183) and following its history from the seventeenth century onwards, one can trace how it intersects with the changing nature of capitalism. Advertising is an institutional method for controlling flows of capital and information; establishing the demands of individuals in order to stabilise an otherwise unpredictable free market; and, beginning in the late nineteenth century, as a way of supporting mass

consumption in highly industrialised societies. Advertising, according to Williams, is not as old as human culture, rather, it is functionally tied to the different stages of capitalism. The Italian Neo-Marxists introduced above argue that capitalism has recently undergone a change of state from Fordism to Post-Fordism. Williams' history of advertising, although first published in 1980, was written in 1961 and thus ends before the explosion of information technologies, globalising tendencies, and restructuring of traditional modes of labour that Post-Fordism describes. Extending Williams' history to cover our current moment helps to contextualise Google's model of advertising historically, as well as the way in which it reflects and co-creates our contemporary form of capitalism.

One of the key narratives of advertising, for Williams, is the expansion of its scope, as its function grew to cover an increasing number of commodities and services. As newspapers grew at the end of the seventeenth century so did the number of advertisements, but only for a specific sort of luxury items or medical quackery: "Ordinary household goods were rarely advertised; people knew where to get these" (172). Modern persuasive advertising, which seeks to establish and perpetuate particular cultural ideals, did not gain dominance until the interwar years of the twentieth-century when it blended with wartime propaganda – posters such as "Daddy, what did you do in the Great War?" (180) – became influenced by advances in modern Psychology, and rode the rising tide of mass media to produce a network of cultural norms that could be bought into through bourgeois products and services. As Williams describes:

> in the 1850s advertising was mainly of a classified kind, in specified parts of the publication. It was still widely felt, in many kinds of trade, that (as a local

> newspaper summarised the argument in 1859) "it is not *respectable*. Advertising is resorted to for the purposes of introducing inferior articles into the market." (173)

It is only in the twentieth century that advertising became the "official art of modern capitalist society" (184), by which Williams means two things. First, it is the aesthetic that covers the walls of our public places, the insides of newspapers, and funds the employment of a whole creative class. Second, it is also "art" in the sense that it relates to advertising as a "magic system": a set of practices and cultural myths to perpetuate an unfulfillable materialist desire that serves as an economic engine. To Williams, advertising should be understood as a kind of grammar for a specific historical moment. Given that Google's dominance in online advertising has led to its parent company, Alphabet, to be valued as the second most valuable company on the Fortune 500, with a market value $579,426 million (Fortune), what can Google's mode of advertising tell us about the grammar of contemporary capitalism online?

It is worth noting that the kinds of traditional mass-market advertisements using slogans, celebrities and jingles that play to our "basic personal relationships and anxieties" (R. Williams 180) are still with us. Many kinds of advertisements that would not be out of place in the context of twentieth-century television or billboards can be found online, from the pre-roll ads of YouTube to the banner ads underneath the masthead of *The New York Times* (see figs. 5 and 6). These advertisements that borrow a familiar form have, however, been incorporated into a different model of how media forms function online. This will be outlined in the second half of this chapter, when the discussion turns to Google's AdSense program. To understand new forms of advertising and their relation to contemporary online capitalism in the

light of Williams' historical narrative, we must pause a while longer on Google's dominant form of advertising: AdWords.

As outlined above, the way in which AdWords functions as a "global real-time, and multilingual market" (Kaplan 59) mapping capital directly to specific words and phrases seems at odds with the kinds of advertising that set to establish a generalised demand in a mass market. The algorithmic rating and auction system that selects a particular sponsored link mean that AdWords provides the most *relevant* advertisement: a listing for an existing demand, rather than a persuasion for something new or different. Google's algorithm, as with its organic results, aims to weed out any misleading, irrelevant or "inferior articles" (as Williams' 1859 newspaper puts it) and, as outlined above, only charges companies for advertisements when, after Google has selected them as the most relevant, they are actively chosen by users. In addition, the standardised format in which sponsored links are presented cuts out the *art* of advertising (see, again, fig. 1). With this outlook, AdWords barely seems like advertising at all. However, these superficial descriptions are not what defines advertising; Google's AdWords functions to structure and control the flow of information and capital in this specific moment of contemporary capitalism. The grammar of digital capitalism is a reflection of Google's structuring of the web. Mass cultural appeal gives way to the long tail of niche commerce; one-way channels of communication and influence become algorithmic feedback loops based around the harvesting of personal data; the growth of immaterial labour expands the reification of previously unmarketable activities into profitable goods and services. AdWords functions as the intermediary form of communication between companies, markets, and individuals that reflects the new grammar of Post-Fordist digital commerce. Google's monopoly on the web is far

from an extended caricature of the hegemony of old-media moguls. Although Google dominates, the system is decentred in a number of ways. The most significant of these is that tracing data patterns as a way of mapping “relevance” draws from the existing behaviours of individuals rather than following any specific normative judgments established by Google. The following section of this chapter introduces several challenges that demonstrate that although Google dominates the online advertising market, control of capital flows is distributed throughout a complex network of users.

3.3 AdWords and the General Intellect

The context outlined above demonstrates the ways in which Google draws information from its users to put to economically instrumental ends. It is in this perspective that Matteo Pasquinelli, in an essay specifically focused on Google’s PageRank algorithm, describes, “Google [as] a parasitic apparatus designed to capture the value produced by the common intelligence” (155). Pasquinelli describes Google as unproductive: seizing the surplus value of already existing networks and establishing a hegemonic power structure that prevents users from accessing the web without Google’s influence. For Pasquinelli, the profits Google makes are part of a wider shift within existing economic and social arrangements, which he describes as “cognitive capitalism” situating his work within a Post-Fordist framework. In doing so, Pasquinelli draws on the work of Antonio Negri, in particular an essay co-authored with Carlo Vercellone in 2007, in which they argue that *rent* serves an important function for current modes of cognitive capitalism, as well as Post-Fordism more widely. Pasquinelli paraphrases their argument:

> rent is the central mechanism of the passage from industrial capitalism to cognitive capitalism. In classical economic theory, rent is distinguished from profit. Rent is the parasitic income an owner can earn just by possessing an asset and is traditionally associated with land property. Profit on the other hand, is meant to be productive and is associated with the power of capital to generate and extract a surplus. (158)

Pasquinelli's criticism is primarily focused on PageRank, Google's algorithm that ranks organic search results for each query. For Pasquinelli, the information that PageRank uses is latent in the network and Google's algorithm is simply organising it rather than creating or producing something new. Google's organic rankings are based on existing patterns of hyperlinks on the web and uses these as an indication of sentiment, much like an academic citation system does. So, although a page with numerous hyperlinks pointing to it (or to continue the citation metaphor, a widely-referenced article) might not have been judged as *good*, it is certainly *relevant* to a particular group of people. Because the main metric of AdWords, relevance, follows the PageRank model, Pasquinelli's argument can be applied to the AdWords mode of advertising. Consequently, as Google's methods of judging relevance are a way of measuring existing behaviours, this information belongs to all users, and as a result, Google is profiting unfairly; in Post-Fordist terminology, Google is renting users their own judgments. For example, searching for [Shakespeare] returns a number of high-quality results (see fig. 4), high-quality in the sense that the results are from reputable sources and pertain to William Shakespeare (of course, judging the results as successful relates to the subjective intentions of a user).

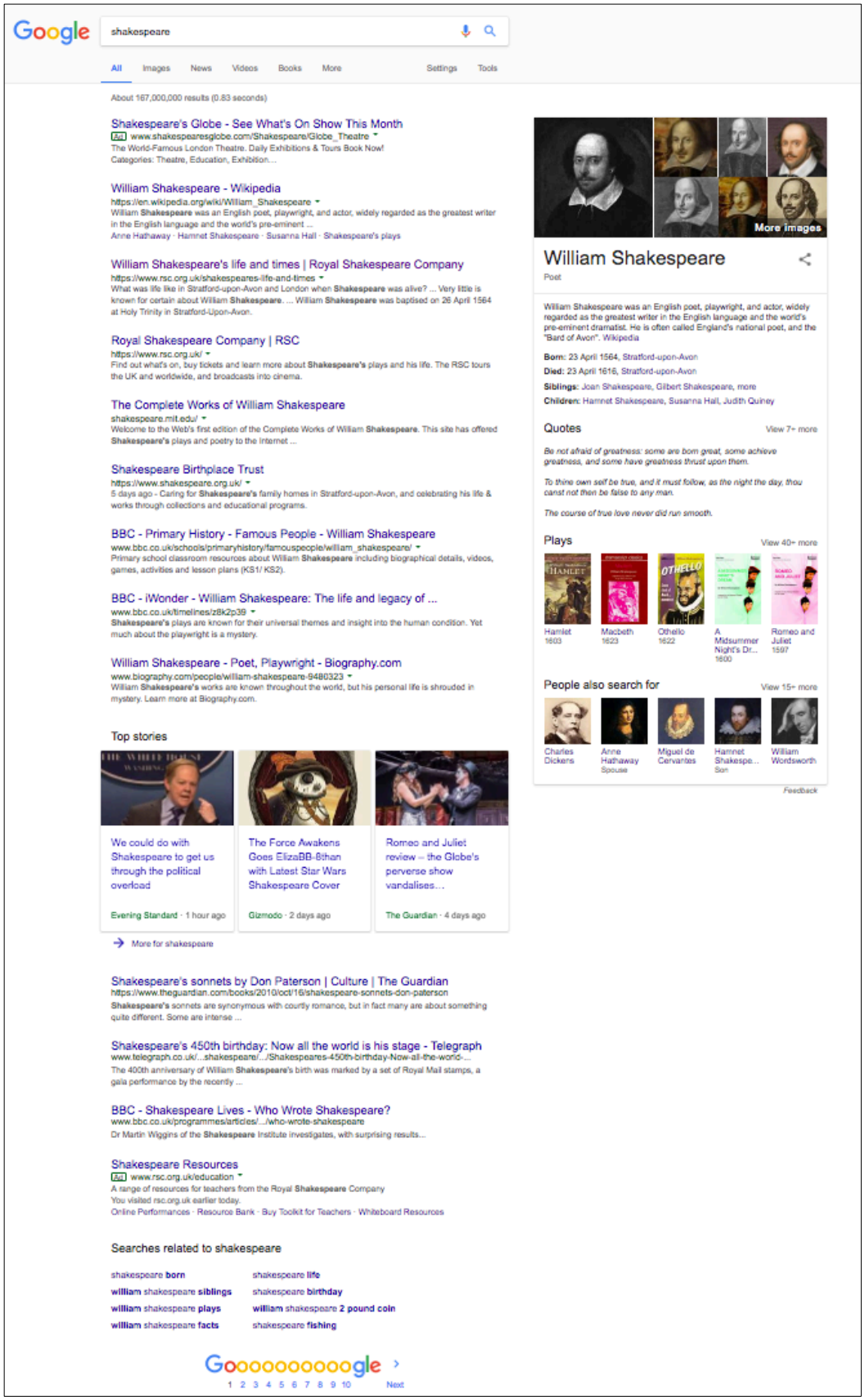


Figure 4. Google results for [Shakespeare] An example of high quality results.

Search performed 7/11/2017.

However, this hierarchy of quality has not been created by Google; the PageRank algorithm simply reflects the already existing values of web users that have linked to these sites. Pasquinelli's description of Google as global rentier is based on the perspective that although we have Google to thank for providing us access to these sites, the much more significant gratitude must go to the digital community of users for creating this curated list through their collective actions online. Therefore, that Google receives revenue for the usefulness of these results is, from Pasquinelli's perspective, wrong; as the responsibility for the curation and hierarchy lies with the community as a whole: journalists, bloggers, and any kind of user that contributes online. These online participants are not getting paid for their contributions. Instead, users are reimbursed through free access to Google's services, regardless of their level of input. This further demonstrates a characteristic of cognitive capitalism in a wider Post-Fordist context: when users are online they are often unknowingly participating in immaterial labour practices and are collectively remunerated through access to a digital service.

Although Pasquinelli argues that "Google itself does not produce any content" (157), rather it profits from the "exploitation of a common cognitive space" (159), it is important to temper this perspective. Search engine functionality is inherently productive; foregrounding the attitudes of users, through whatever means, is a necessary part of building a robust infrastructure for the web. However, the influence that Google have over the web means that although the web might be a collectively produced space, we have essentially placed its ownership in private hands. The terms of this debate are central to an understanding of Google's role online and while the terminology borrowed from the history of Marxist critique can be helpful, a

critique of current technology must be wary of anachronistic deployments that disfigure current forms of capitalism into modes of pre-digital industrial capitalism.

To describe Google's advertising with Post-Fordist terminology we can say that AdWords is a way of functionalising the "general intellect". Here, Paolo Virno's reading of this term and his way of updating Marx's original meaning is important:

> Marx, without reserve, equated the general intellect (that is, knowledge as principal productive force) with fixed capital, with the "objective scientific capacity" inherent in the system of machines. In this way he omitted the dimension, absolutely preeminent today, in which the general intellect presents itself as living labor […] In the Post-Fordist environment, a decisive role is played by the infinite variety of concepts and logical schemes which cannot ever be set within fixed capital, being inseparable from the reiteration of a plurality of living subjects. The general intellect includes, thus, formal and informal knowledge, imagination, ethical propensities, mindsets, and "linguistic games". […] The general intellect becomes an attribute of living labor when the activity of the latter consists increasingly of linguistic services. (106)

An important feature of the general intellect, as described here, is that it cannot be reduced to simply what a collective *has* produced, but what it is *capable* of: the shared "*faculty* of thinking; potential as such, not its countless particular realisations" (66). As contemporary capitalism continues its trend towards Post-Fordist relations that are underpinned by cognitive or immaterial labour, the general intellect can be functionalised like never before. Google's financial profits are far

from the only benefits of the web, but allowing them to be consolidated reduces the collective control of the general intellect, the “formal and informal knowledge, imagination, ethical propensities, mindsets, and ‘linguistic games’”, that establish and sustain community.

This situation of privately regulated hyperconnection results in what Virno calls “a publicness without a public sphere” (36). The web has actualised the previously theoretical connectedness of the multitude, however, the web has not become a democratically shared space: instead, it is owned and structured around results that bring economic value and valorise non-economic ideas. Rather than generalising about the web as one enormous public sphere, it is more accurate to describe the web as a congregation of various multitudes. Each of these publics have different levels of visibility and influence. Importantly, Google AdWords has an uneven coverage of these different multitudes and this has a significant impact on the shape and scope of the web.

Google subdivides users and delivers different results based on a large number of criteria (over 200), such as language, geographical location, and previous search history. The weighting of each criterion changes and is re-established every time a query is searched; it is impossible for users to know which metrics and characteristics influenced their results. Each search causes users to be drawn into a particular multitude, each with its own general intellect and each with its own corresponding sets of search results. The logic of advertising then forces particular criteria to have a disproportionate influence. Jeff Huber, Head of Ad Engineering at Google, when interviewed by Levy underlined the importance of national space in advertising. Quoted in *In the Plex*, Huber says:

> Yes, search is a huge system, but it's stateless — you can easily serve it from ten different places in the world, and if this version is slightly different than that version, the user won't know, nobody will notice. But with advertising, the state is important, because advertisers are always updating their campaigns, and micro transactions are happening at ferocious rates per second, and all that has to be synchronized. (116)

Therefore, because of the advertising incentives, a particular metric such as location gains a more significant influence in the kinds of results provided. In addition, the time and effort spent on engineering results follows from economic motivations, prioritising criteria that correspond to specific users. This leads to a situation in which some languages become more valuable to Google than others. In turn, because some languages have less economic value on the web, the multitudes associated with those languages are accommodated for far less.

There is a lot at stake in debates concerning the ownership of information and good Internet citizenship requires that we stay attentive to Google's role; however, the general intellect, as the current potential for thought, is far more than a collection of specific documents that may or may not be used to extract surplus value. Describing Google as merely parasitic ignores the productive capacities that Google enables by making the web usable. The accessibility of the web that Google's search engine has enabled has allowed for emergent activities and communities that might not exist otherwise. In addition, Google *does* enable financial remuneration to many online content creators through its second advertising program, AdSense, the benefits and drawbacks of which are outlined in the second half of this chapter. Nevertheless, it is important to observe that these are two separate systems and the

beneficiaries of AdSense are by no means the people who contributed to the structure of the web that AdWords profits from. An understanding of AdSense, provided in the latter part of this chapter, in many ways problematises Pasquinelli's distinction of Google as the global rentier. However, even though AdSense functions very differently than AdWords, an analysis of the two systems provides further evidence as to why Post-Fordist relations are a key lens for understanding Google Search *and* Google as a company.

3.4 The Future of Languages in an Incentivised Online Environment

Before moving on to a discussion of AdSense, which can be described as a way *individuals* are compensated, or in less generous terms, how particular profitable behaviours are encouraged while others are deincentivised, we must outline a specific consequence of AdWords: its effect on languages. The latent knowledge contained within a language is a key part of the general intellect and a central theme in Post-Fordist descriptions of contemporary capitalism. Virno uses Gilbert Simondon's concept of individuation to describe how a language structures a multitude: "Language is pre-individual; it is the historical-natural language shared by all speakers of a certain community. Language belongs to everybody and nobody […] the use of the spoken word is, at first, something inner-psychic, social, public. A 'private language' does not exist" (77). Each query entered into a search engine is part of this individuating process: an interaction with a wider multitude that through language constructs an individual as a subject. Subjects cannot stand apart from their shared language, as Virno paraphrases Simondon to argue that "individuation is never concluded […] the subject consists of the permanent interweaving of pre-

individual elements and individuated characteristics; moreover, the subject is this interweaving" (78). It is, therefore, important for the very subjects that are continually emerging that language does not become wholly co-opted by one single economic description. In particular, the reification of AdWords means that the future of a number of languages is at stake, alongside the multitudes that speak with them.

The lack of diversity of language online is profound. Daniel Prado outlines that: "Barely 5% of the world's languages have a presence in cyberspace" (34). The decline in language diversity is a general global trend off- and online and Google's impact on the web has followed, if not exacerbated, these general trends. "The Globalization Group (2010) suggests that 90% of total international GDP is produced by the speakers of only 14 languages" (38). Google have no incentive to provide their services to language groups that do not represent a profitable market. In addition, the populations that are underrepresented on the web often correspond to less economically developed areas, due to the infrastructure and costs required to gain an online presence. Many of these individuals already have a choice between more than one language, as Viola Krebs and Vincent Climent-Ferrando attest in "Languages, Cyberspace, Migrations": "It is estimated that close to one half of the world's population is bilingual" (232). However, the usefulness of a second language online may contribute to a deterioration of a more localised primary language, and the culture that it is tied to. As Prado notes, when communities from less developed countries come online, they choose not to use their native language: "A 2003 study by Marcel Diki-Kidiri showed that in a sample of 1,374 African sites, only 3.22 % used an African language as the language of communication" (39-40). This creates a feedback loop in which, the more that new users find their own language underrepresented, the less likely they are to use it. Therefore, the notion that new

web users ever have a choice to start with is radically limited. The web has facilitated the growth of economically useful languages but has prevented a number of languages from ascending to the web. In his 2013 article, Andras Kornai argues that this trend has developed significantly enough to be considered irreparable: the findings of Kornai's team was that "the vast majority (over 95%) of languages have already lost the capacity to ascend digitally" (2). Languages can disappear online if there is simply one usable alternative. This feedback loop, that limits linguistic diversity, is perpetuated by Google's financial model. The online success of some languages and the failure of others is, in a sense, payment from Google to particular linguistic communities. If users create content in a specific language, Google can harvest its data and enable advertising in that language, which makes searching in that language profitable for Google. If no one is using a language there is no incentive for advertisers to pay Google for specific words and phrases, thus Google accelerates the process of online language death. Therefore, the control over which languages survive is only partially the responsibility of content creators and users; the power is predominately given over to Google.

This process, in which social responsibility is given over to a private company, such as Google, is what Vaidhyanathan describes as "public failure" (6). Google have succeeded in dominating many aspects of people's digital and embodied lives worldwide; many of Alphabets' enterprises make significant losses and are backed up by Google's large advertising revenue; competitors without such a large revenue stream have thus been overtaken or been bought up by Google. Therefore, a discussion regarding Google's financial success is also a discussion of a deal that the global public have made with one company. As Siva Vaidhyanathan puts it:

> Because of its ease and power, because it does things so cheaply and conveniently, it may cause us to miss opportunities to do things better. Google's presence in certain markets, such as advertising or book search, retards innovation and investment by potential competitors because no one can realistically wrest attention or investment from Google. And when Google does something adequately and relatedly cheaply in the service of the public, public institutions are relieved of the pressure to perform their tasks well. This is an important and troubling phenomenon I call *public failure*. (6)

In the above quotation, Vaidhyanathan is focused primarily on Google's book scanning project and way in which governments and universities have allowed Google overwhelming control over the future of digital textual content – digitised analog content as well as born-digital text – because Google covers the costs of investment. Vaidhyanathan's term, "public failure", is equally applicable in the context of Google's advertising empire. Google is shaping the world we live in by dominating markets through advertising revenue which, in turn, creates more advertising opportunities. Free services, like Google Search, are the payments back to specific multitudes in exchange for using their general intellect for profit. Considering the profits involved, it could be argued that the public should be considered as employees of a newly Post-Fordist workforce, creating financially profitable data-sets in their "spare time", similar to the way that Uber drivers make additional revenue in their "spare time".[14] Also, considering the job losses occurring worldwide (due, in part, to the changing nature of work) this wealth needs to be better shared. However, these arguments have revolved primarily around the taxes

[14] Research has shown that, for most people working in flexible labour roles, this work is not their primary means of income. See, for example, Hall and Krueger, who show that 51% of Uber drivers "drive for less than 15 hours a week, and […] 85 percent chose to drive less than 35 hours a week".

that Google, and many other multinational companies, avoid paying.[15] However, to see Google's profits within a Post-Fordist context, and to describe its users as quasi-employees producing a linguistic landscape for Google to profit from, the payments back to the multitudes should be on a different scale from the current tax systems. Google is not simply making money from the creativity of individuals, but rather shaping all cultural experience on the web into a system that can be easily reified and commodified.

3.4.1 Google's International Expansion

A significant part of Alphabet's agenda, as a company, is to extend their reach and increase the number of contexts that they can transform into profitable ventures. An example of this is their drive to make as much of offline life machine readable, as mentioned earlier in reference to Google Glass and its related patents. Another kind of expansion that Alphabet has invested in is their drive to provide Internet access across the globe. Google's Project Loon, one of the Alphabet subsidiaries developed using Google's advertising profits, is a "network of balloons travelling on the edge of space, designed to extend Internet connectivity to people in rural and remote areas worldwide" ("Google: Project Loon"). The project has been described by the company as a kind of social mission, extending the Internet to 4.3 billion people, but in the *MIT Technology Review* Tom Simonite writes:

> It is odd for a large public company to build out infrastructure aimed at helping

[15] Google avoids taxes in a number of different ways, including housing profits in Bermuda, to which Eric Schmidt commented "I am very proud of the structure that we set up. We did it based on the incentives that the governments offered us to operate" (qtd. in Kavoussi).

> the world's poorest people. But in addition to Google's professed desires to help the world, the economics of ad-supported Web businesses give the company other reasons to think big. It's hard to find new customers in Internet markets such as the United States. Getting billions more people online would provide a valuable new supply of eyeballs and personal data for ad targeting. That's one reason Project Loon will have competition: in 2014 Facebook bought a company that makes solar-powered drones so it can start its own airborne Internet project.

In this way, it describes another "public failure" in which a "blessing" becomes a "necessary – seemingly natural – part of our daily lives" (Vaidhyanathan 6-7). This provides opportunities for the greater logic of Post-Fordism to enter into new rural contexts such as "isolated parts of Brazil, Australia, and New Zealand" (Simonite). Simonite cites Sunil Abraham, executive director of the Centre for Internet and Society, a think tank in Bangalore, writing that, Abraham

> is also wary of Project Loon because of the way Google and other Western Internet companies have operated in developing countries in recent years. They have cut deals with telecoms in India and other countries to make it free to access their websites, disadvantaging local competitors. "Anyone coming with deep pockets and new technology I would welcome," he says, but he adds that governments should fix up their patchy regulatory regimes first to ensure that everyone – not just Google and its partners – really does benefit. (qtd. in Simonite)

In this way, not only do the profits of advertising come to underlie the epistemological landscape of the web, but also the infrastructure required for Internet access, offered for free to individuals in developing countries. The potential benefits of opening the Internet up to a wider range of individuals, from different cultures, speaking different languages are multifarious; however, these benefits are not inevitable. If we allow Google, or any other single company, to dictate the physical and digital infrastructure of the web, those multifarious benefits, in order to gain traction, will have to conform to the particular dominant economic affordances of the web. Given the impact of AdWords' reification of language upon linguistic diversity, outlined above, such international expansion may only solidify the current boundaries and limitations of the web. This is not to say that Google's actions are implemented in bad faith, only that the hegemonic situation means that any diversity that is not profitable will struggle and the benefits of widening our online community may well be drastically limited as a consequence. It is in this context, of economic affordances dictating the landscape of the web, that we turn to AdSense, the other side of Google's advertising coin.

4.0 AdSense and Post-Fordism

AdSense is the second, smaller, advertising programme that Google operates. AdSense is Google's brokerage programme that enables online content creators to monetise their content by placing third-party adverts on their websites, blogs, or YouTube videos. Sullivan uses the analogy that AdSense "basically turned the Web into a giant Google billboard. It effectively meant that Google could turn everyone's content into a place for Google ads" (qtd. in Auletta 91). AdSense allows users to monetise their online content by setting aside spaces (see figs. 5 and 6 for an example) that Google can fill with adverts that are relevant to the content of the site and/or the user visiting the site.

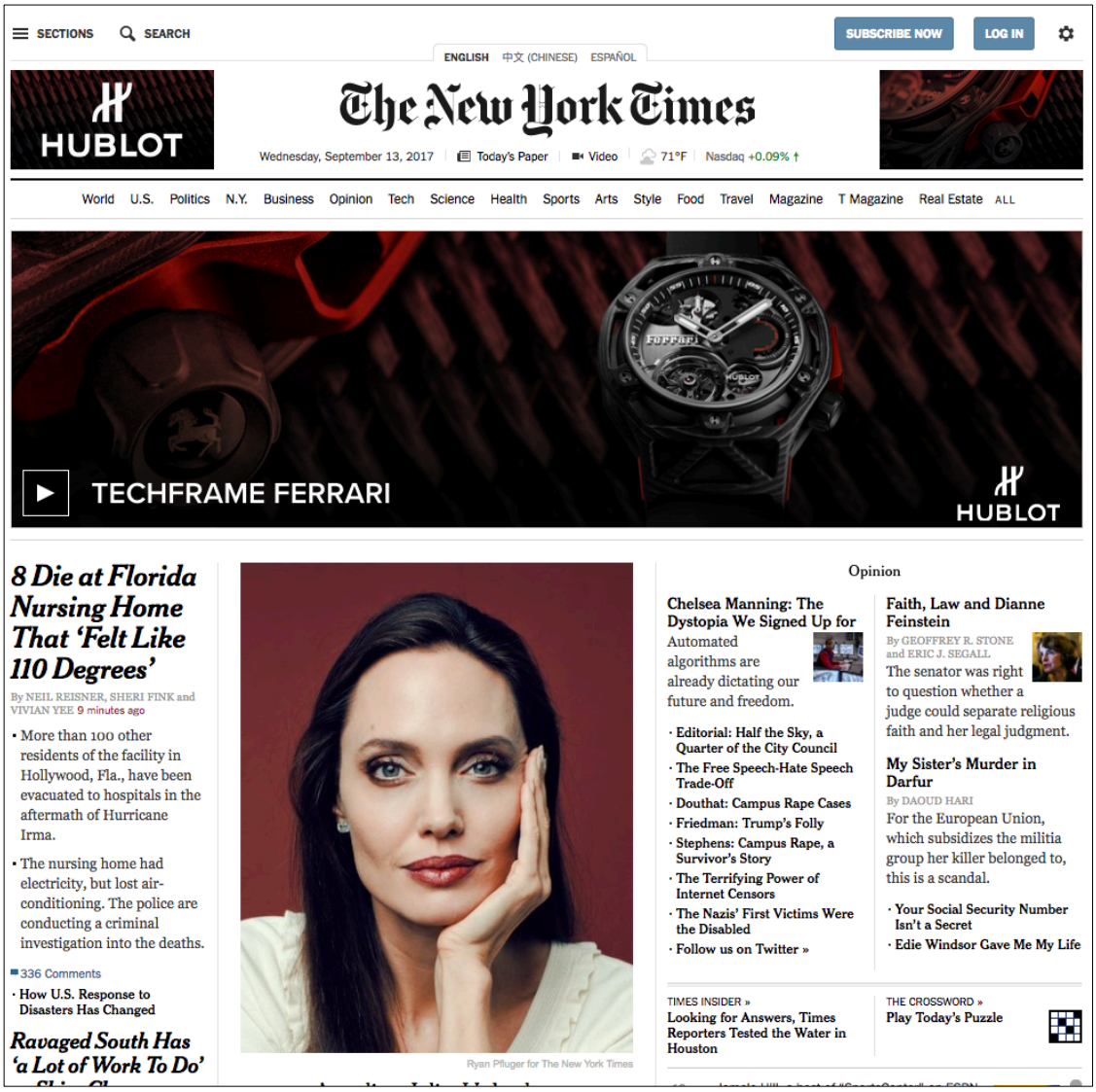


Figure 5: Examples of AdSense banners on the front page of *The New York Times* website. Accessed 13/09/2017.

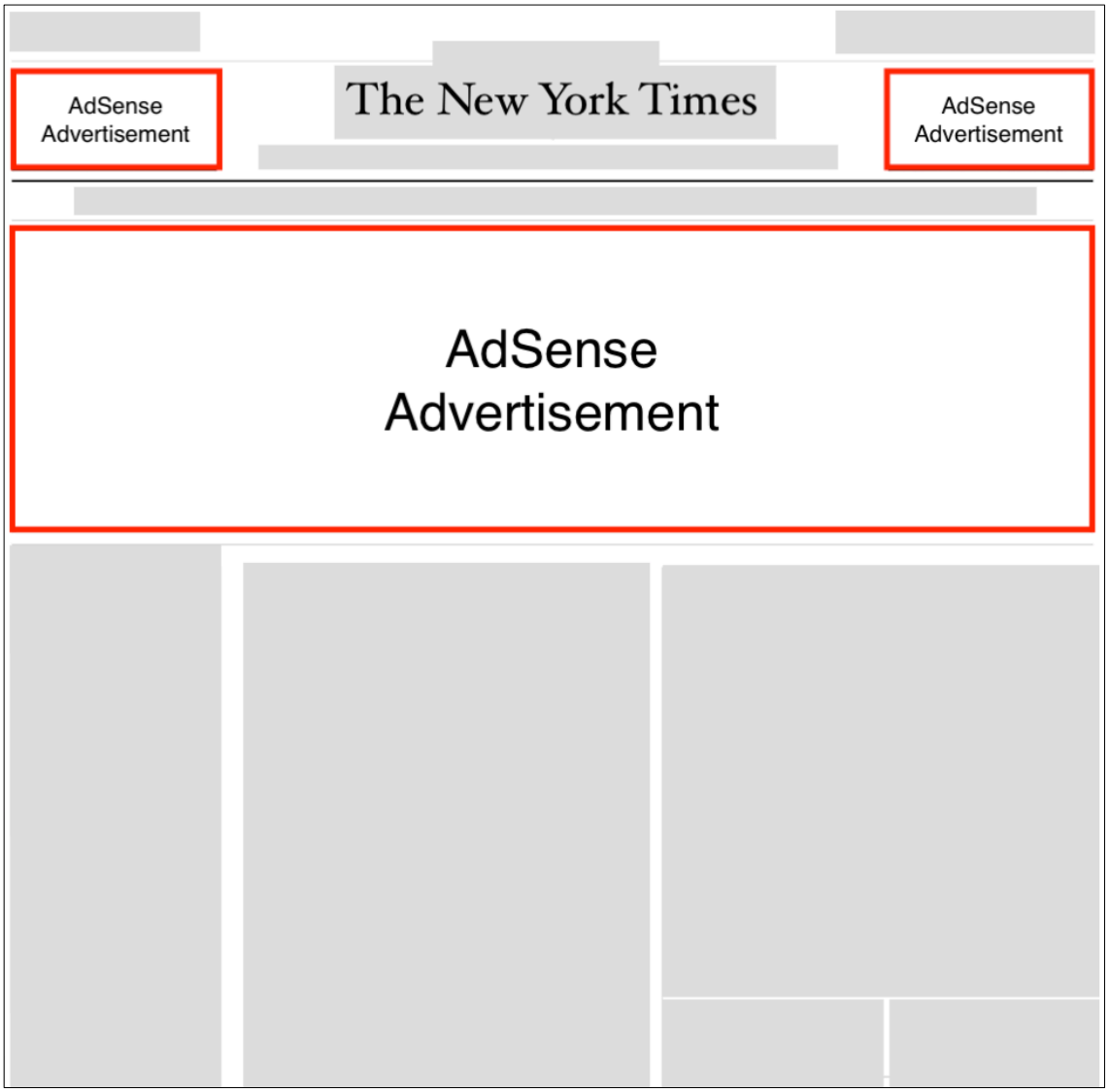


Figure 6. A simplified layout of figure 5 highlight which parts of the page are AdSense banners. Drawn by the author.

This is accomplished in the following way:

> Google's software crawls the site, performs semantic analysis on the text on each page, and then automatically selects ads that are displayed […] matched to the meaning of the text. It calls this "contextual advertising." Site owners and Google split the proceeds when visitors click on the ads. (Stross 159)

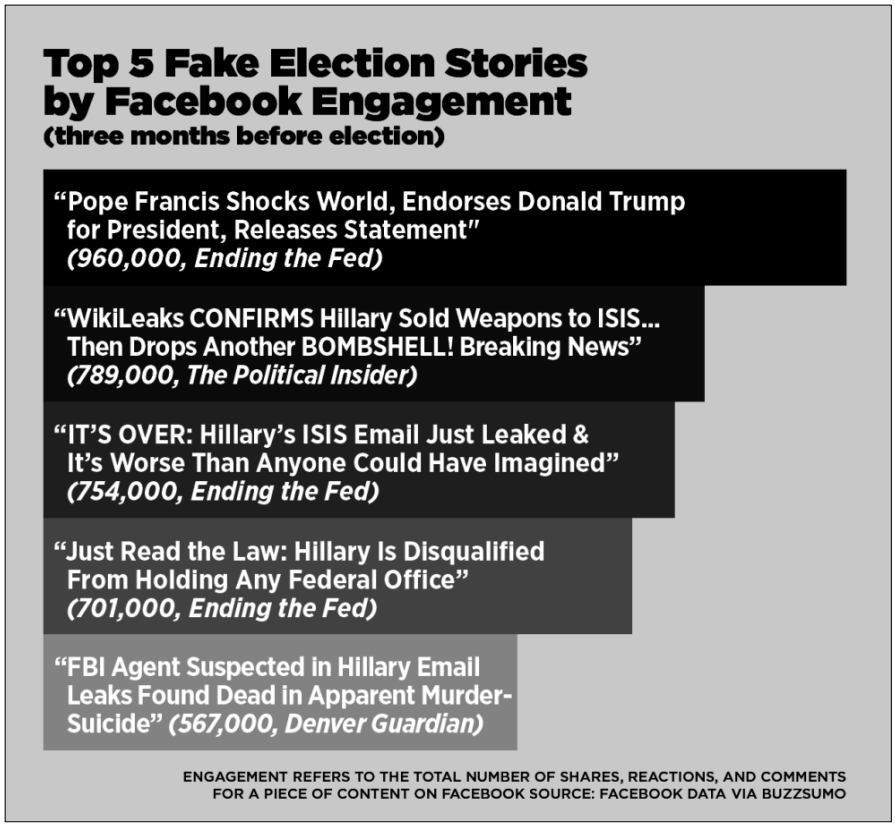
This process of matching advertisements to content and users has had a large impact on the landscape of the web today, reinforcing associations and shaping what kinds of activities can become profitable. In addition, AdSense as the most dominant online advertising provider has shaped the flows of information and capital in a way that promotes Post-Fordist labour relations between Internet users. There are many benefits to the AdSense model and without it the web would be unrecognisable. However, there are serious negative consequences to a system that values page views, engagements, and shares above all else. This section will cover the impact of AdSense on online discourse and, in particular, analyse the profitability of fake news in the 2016 US presidential election.

4.1 AdSense and Fake News

It is important to have a clear definition of fake news because it refers to a specific phenomenon that grew in influence during the 2016 US presidential election. Since entering office, President Trump has adopted the phrase, using it against legitimate media organisations, such as CNN, in a way that has muddied the term. Here, the term fake news is used in line with Hunt Allcott and Matthew Gentzkow's definition as "news articles that are intentionally and verifiably false, and could mislead readers" (213). In this way, fake news refers to a specific discourse that proliferated around the presidential election and documented through investigations by *BuzzFeed* News, *The Guardian*, and in Allcott and Gentzkow's article "Social Media and Fake News in the 2016 Election". Of particular importance is that although the most prominent topic of fake news was – and still is – politics, analysis suggests that

these fabricated news stories were written purposely for viral impact and often created purely for profit. Their success can be used to outline some media-specific characteristics of the web and the consequences of structuring the online news environment around advertising revenue. Fake news is not new, nor is it without precedent; Allcott and Gentzkow's article, currently the only academic paper that specifically addresses fake news, provides a multi-disciplinary literature review concerning related phenomena, such as conspiracy theories, as well as a survey of journalistic reports that give examples of individuals who have produced fake news for a number of years. However, the increased proliferation and potential impact of fake news in the 2016 presidential election represents a sea-change, one in which Google's AdSense programme and Google's relationship to the second biggest online advertiser, Facebook, looms large.

Although it is not unusual for the US presidential elections to become an axis around which false claims circulate, the 2016 election saw a different level of misinformation. Articles with inflammatory titles, and which could easily be debunked, such as: "Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement", "Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America", and "Van Full Of Illegals Shows Up To Vote Clinton At SIX Polling Places, Still Think Voter Fraud Is A Myth?" circulated on social media, in particular on Facebook, at an unprecedented scale (see fig. 7 for a comparison of hoax to mainstream news engagements).



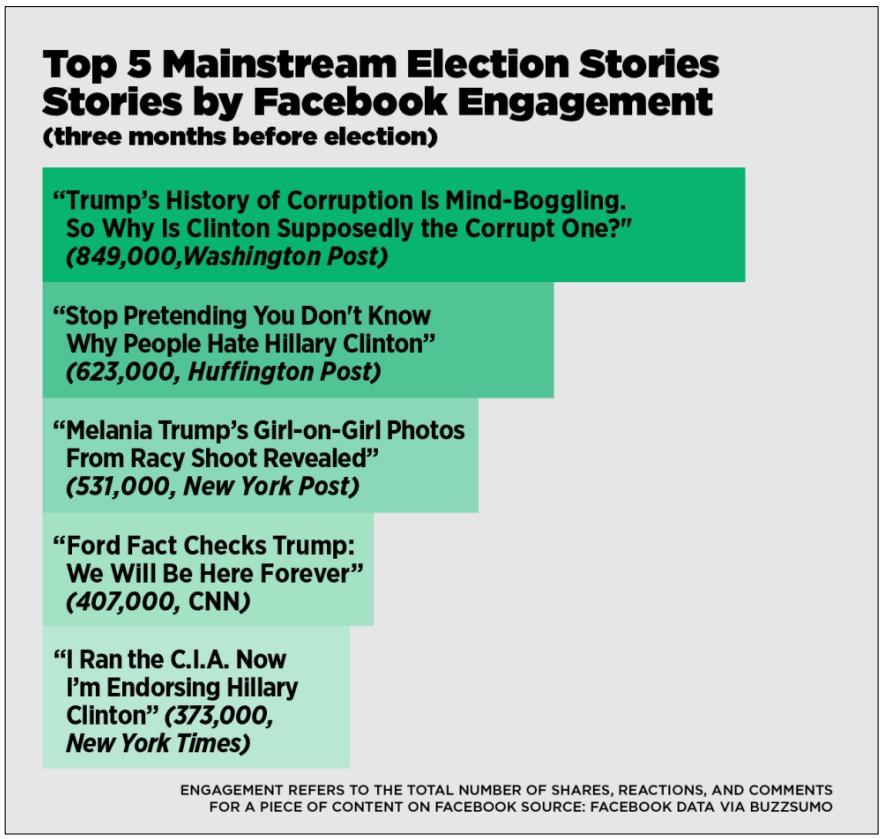


Figure 7. Fake vs. mainstream news stories, ranked by Facebook Engagements.

Provided in Silverman "This Analysis Shows".

In fact, analysis by *Buzzfeed* News showed that in "the final three months of the US presidential campaign, the top-performing fake election news stories on Facebook generated more engagement than the top stories from major news outlets such as *The New York Times*, *Washington Post*, *Huffington Post*, NBC News, and others" (Silverman). Although some of the sources that created and disseminated these stories did so for politically motivated reasons, a number of investigations uncovered that the majority of fake news stories were engineered purely to make money from online advertising networks, particularly Google's, and designed without any political motive. It was reported by *BuzzFeed* and *The Guardian* that

> Over the past year, the Macedonian town of Veles (population 45,000) has experienced a digital gold rush as locals launched at least 140 US politics websites. These sites have American-sounding domain names such as WorldPoliticus.com, TrumpVision365.com, USConservativeToday.com, DonaldTrumpNews.co, and USADailyPolitics.com. They almost all publish aggressively pro-Trump content aimed at conservatives and Trump supporters in the US. […] The young Macedonians who run these sites say they don't care about Donald Trump […] These sites open a window into the economic incentives behind producing misinformation specifically for the wealthiest advertising markets and specifically for Facebook, the world's largest social network, as well as within online advertising networks such as Google AdSense. (Silverman and Alexander)

Although not all the hyperpartisan, clickbait, and hoax news came from Veles – many news reports also confirmed sources in the US – the geographical and political

detachment of those working from Macedonia exemplified the culture of making money through advertising revenue regardless of the content or consequences. Hannah Parkinson's report in the *The Guardian* described the Veles fake news writers as young people who were simply "non-partisan kids looking for cash just catering to demand". Many of the individuals producing fake news about the election had already been earning a good living from advertising revenue of websites aimed at an American audience. Many of the Macedonian creators agreed to be interviewed and it becomes clear that, although a small minority favoured Trump, the subject matter chosen was interchangeable and based around profitable topics. Many of the writers had been creating viral content for years, for example:

> In Veles, Aleksandar and Borce Velkovski are so renowned for the health food website they started that they're known as the Healthy Brothers. HealthyFoodHouse.com is a jumble of diet and beauty advice, natural remedies, and other nostrums. It gorges on advertising as it counsels readers to put a bar of soap under their bedsheets to relieve nightly leg cramps or to improve their red-blood-cell count with homemade beet syrup. Somehow the website's Facebook page has drawn 2 million followers; more than 10 million unique visitors come to HealthyFoodHouse.com every month. (Subramanian)

The creators reinvest part of their earnings to buy fake Facebook profiles and by paying Facebook directly to promote their pages. The analytic feedback tools provided by Facebook and Google allow the creators to develop a good understanding of the criteria that drive content to become viral. Although politically focused content is a recent adaptation, the culture of seeing this kind of activity as a

sustainable job had been established for years in Veles. Another creator from Veles, Mirko Ceselkoski, explained in an interview with Samanth Subramanian that he began

> in the early 2000s. He built seven or eight websites—about muscle cars or celebrities or superyachts, all oriented toward the American reader, because an American reader is roughly three times more valuable than a non-American one. For five or six hours of daily toil, Ceselkoski says, you can earn approximately $1,000 a month. Many Macedonians can spare the time; the unemployment rate is around 24 percent.

In 2011 Ceselkoski began teaching courses in creating viral media and coaching other individuals, for significant tuition fees, in viral content creation and ways of driving Facebook engagements. Those who took classes from Ceselkoski included the Healthy Brothers as well as individuals responsible for pro-Trump fake news sites. The previous profits pale in comparison to the success of election fake news "Between August and November, Boris [18 year old from Veles, real name redacted] earned nearly $16,000 off his two pro-Trump websites. The average monthly salary in Macedonia is $371" (Subramanian). The nature of this Macedonian cottage industry bears some similarity to various international cyber-crimes, such as the 419 advanced fee scams including the infamous Nigerian prince scam. The fake news writers might not be directly taking money from individuals but they are making money from misleading those in more economically affluent countries. Fake news is not considered criminal in the way that 419 scams are; however, both activities articulate similar characteristics of a globalised online space.

This kind of relationship between individuals and institutions operationalised through international flows of information, labour, and capital exemplifies a Post-Fordist logic outlined at the start of this chapter. Fabricated stories or not, all creation of online content, whether news or opinion, is a form of precarious employment that foregrounds various aspects of Post-Fordism, many of which lie outside the scope of this chapter. If this chapter were to focus directly on the kind of employment of those writing fake news stories, a Post-Fordist lens would highlight the immaterial, flexible, individualised nature of that work. However, it should be clear that the proliferation of fake news impacted and incorporated a much larger group of individuals than simply those writing the stories. By focusing on the wider effects, rather than specifically the niche group of writers, the consequences of online advertising supported content demonstrates a proliferation of Post-Fordist logic, even in the lives of those whose employment is firmly material, industrial, or Fordist.

In particular, the writing, reading and sharing of fake news articles can be seen as components of an immaterial labour arrangement. Immaterial labour is described by Maurizio Lazzarato as

> the activity that produces the “cultural content” of the commodity, immaterial labor involves a series of activities that are not normally recognized as “work” — in other words, the kinds of activities involved in defining and fixing cultural and artistic standards, fashions, tastes, consumer norms, and, more strategically, public opinion. (133)

Immaterial labour is not unique to Post-Fordism, it has served a role stabilising previous economic arrangements, for example, the marketing of cultural values in

the twentieth-century that Adorno and Horkheimer refer to as the culture industry. However, the immaterial labour of producing fake news does not shape the values of consumers in order to sell them commodities. Instead, fake news functions by both fixing certain norms and acting as the commodity that matches such a demand; however, such a demand is exacerbated, rather than fulfilled. These norms and values are part of another element of Post-Fordist labour: affective capitalism. Such a framework highlights the affective qualities of fake news: the reactions it creates as well as the social atmosphere it builds upon. Kylie Jarrett, following the work of Brian Massumi, describes affect as

> those sensory experiences of movement and feeling that are part of the social, cultural and psychological experience of individuals, but which lie beyond the directly signifying properties of discourse. It is differentiated from emotion for it is as an embodied intensity that is outside conscious articulation. (*Feminism, Labour and Digital Media* 121)

Here, the similarity to the notion of the "general intellect" should be noted. Jarrett uses such a perspective within a feminist critique of digital labour in order to argue that it is a mistake to describe individuals online only in terms of rational action. Isto Huvila continues such a call in the context of search engines. Huvila argues that rather than seeing individuals as rational information seekers, what "counts in the contemporary affective economy of knowing, is the affective attachment to a sensation of being able to know" (577). However, although affect must have a significant bearing on the spread of fake news, it is difficult to establish a clear understanding of how and why fake news gained traction, with readers, in the way

that it did. Were individuals convinced of the accuracy of the stories or were they shared for other reasons? In taking a large-scale economic focus, this chapter outlines the economic and advertising-dependent incentives behind the creation and proliferation of fake news, but does not address the smaller-scale actions of individuals, rational or otherwise, on which such a proliferation depends. However, even if nothing is known about their reception, it is the economic structure that provided the informational affordances required for fake news to flourish. It is this economic structure that was jointly reinforced by the way in which Google and Facebook prioritise online content.

4.2 The Reciprocal Relationship Between AdSense and Facebook

Although this international aspect of fake news is important, many of the most successful fake news creators reside in the US, for example, *Liberty Writers News* who were the focus of an interview by *The Guardian*:

> Liberty Writers News, a two-person site operating out of a house in the San Francisco Bay Area, generates income of between $10,000 and $40,000 a month from banks of ads that run along the side and bottom of every story. Paris Wade and his partner Ben Goldman have mastered the art of getting traffic. The ability to write a clickbaity headline, toss in some user-generated video found on YouTube, and dash off a 400-word post in 15 to 30 minutes is a skill they don't teach in journalism school, says Wade, who graduated from the University of Tennessee with a degree in advertising. (Tynan)

The banks of advertising described were indicated as those provided by AdSense (SadBotTrue), although interviewees described using multiple platforms. The pair disclose that they "spend around $3,000 a month paying Facebook to promote the page" and that "95% of our [*Liberty Writers News*] traffic is coming from Facebook". Here, Facebook benefits in two ways. First, *Liberty Writers News* pay Facebook directly to promote their site and second, *Liberty Writers News* urge their readers to "Share this right now! Let's beat the liberal media to it. Share, share, share it all over Facebook" (Tynan), which in turn increases the time users spend on Facebook and provides further advertising opportunities. Facebook and Google have little incentive to stop a spread of misinformation as it represents some of the most profitable content on which they host advertisements. In an interview with *Bangor Daily News*, one US-based fake news creator, Paul Horner, explained his involvement:

> My sites were picked up by Trump supporters all the time. I think Trump is in the White House because of me. His followers don't fact-check anything – they'll post everything, believe anything. His campaign manager posted my story about a protester getting paid $3,500 as fact. Like, I made that up. I posted a fake ad on Craigslist. (qtd. in Dewey)

In addition to Trump's campaign manager, Corey Lewandowski, circulating and promoting Horner's fake news, Caitlin Dewey reports that Eric Trump and Kellyanne Conway also tweeted fake news stories from Horner's site. The original version of the *Breitbart* article "Palin on Paid Anti-Trump Protesters: 'Not Even President Yet and Our Guy's Already Creating Jobs'" (Moons) cited Horner's fake news story as a source. Horner's fake news site, abcnews.com.co, still hosts the original article

"Donald Trump Protester Speaks Out: 'I Was Paid $3,500 To Protest Trump's Rally'" but now starts with an added message that "this story is not real […] This story is mocking all of you sheep who think protesters are getting paid" (see fig. 8). Searching using the Internet Archive's Wayback Machine we can see that this message was not part of the original story (see fig. 9).

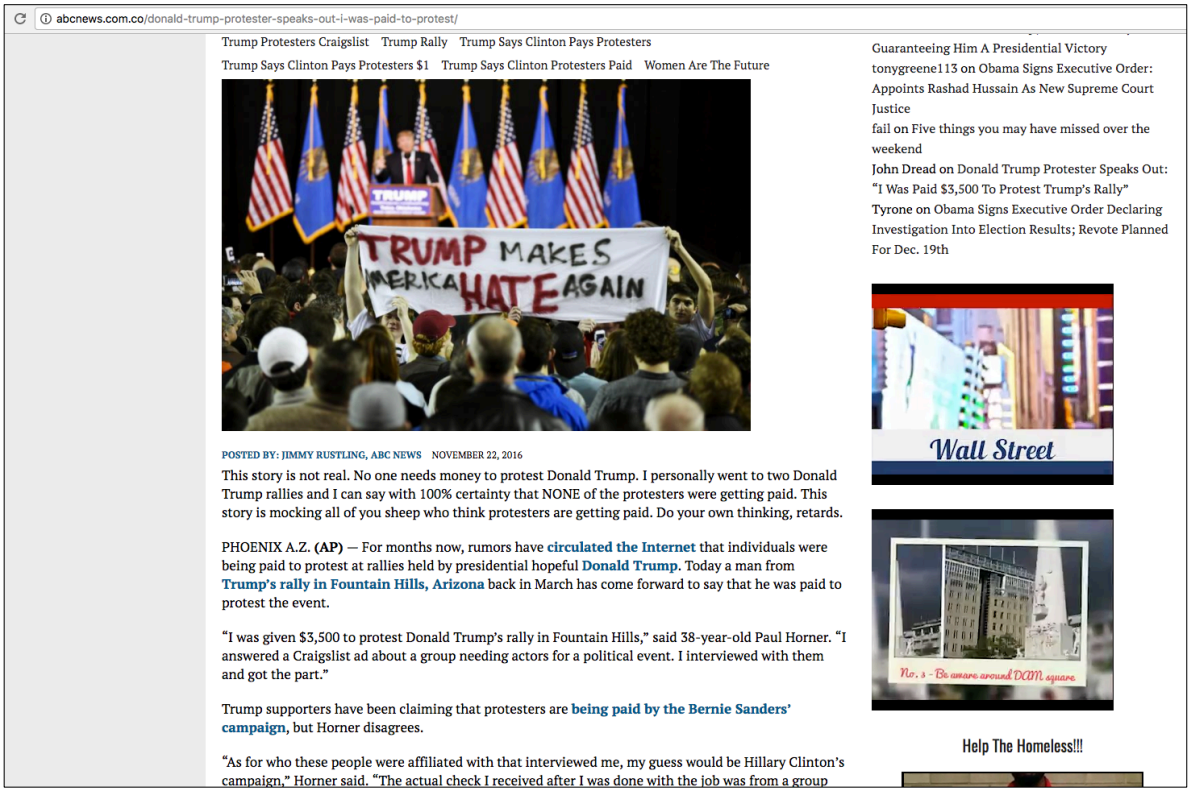


Figure 8. Paul Horner's "Donald Trump Protester Speaks Out: 'I Was Paid $3,500 To Protest Trumps Rally'" article, screenshot as of 8/06/17, including "This story is not real" paragraph.

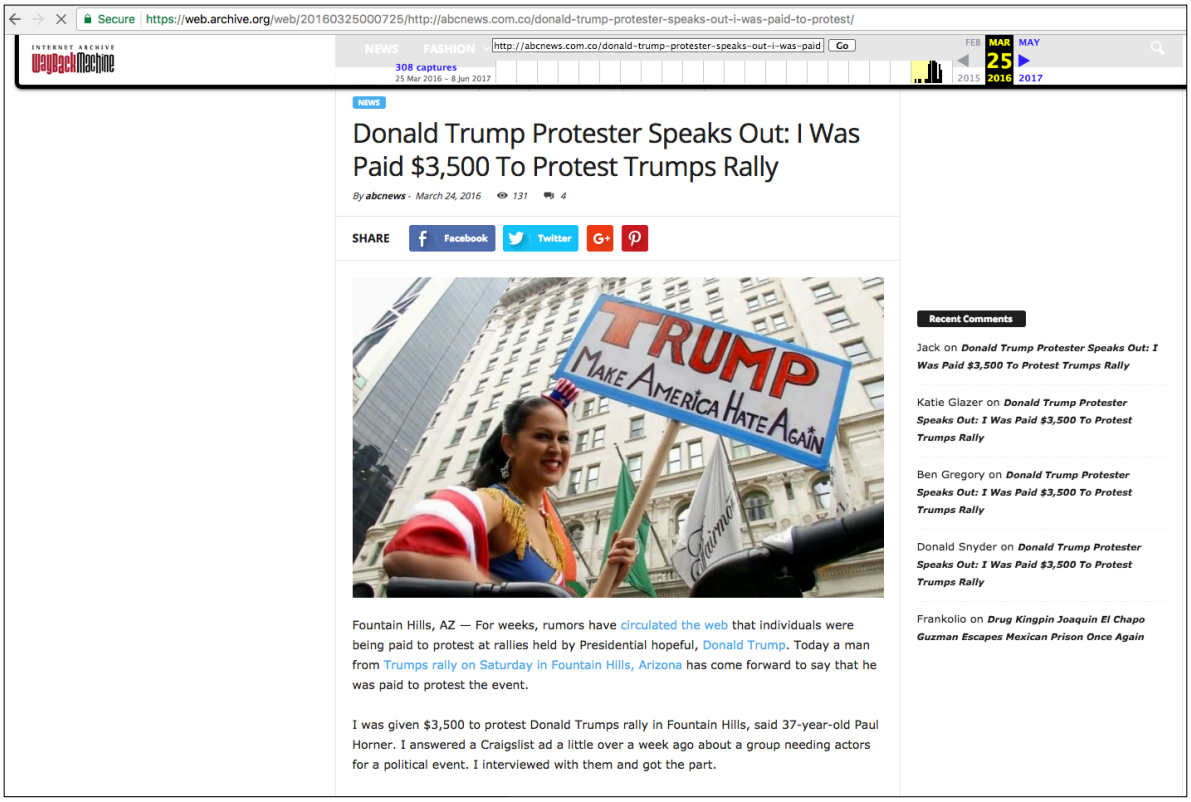


Figure 9. Paul Horner's "Donald Trump Protester Speaks Out: 'I Was Paid $3,500 To Protest Trumps Rally'" article, as of 25/03/16 (via the Internet Archive's Wayback Machine), without the "This story is not real" paragraph.

Since the results of the election Horner has taken part in numbers of interviews with established news organisations, Channel 4 for example, in which he has promoted himself as someone trying to make the public better informed. In the context of digital capitalism, whether his articles were meant to fuel or satirise Trump supporters is unimportant. Horner, along with many others, has financially benefited from flows of information that are promoted and incentivised by Google AdSense and social media sites such as Facebook.

As far as advertising revenue is concerned, Google have no incentive to care whether these stories are being read as satire by Democrats or sincerely by Republicans. Dewey's article describes how Horner has "made his living off viral

news hoaxes for several years". In response to a question referring to the announcements from Facebook and Google that they would stop allowing fake news sites to use their advertising services, Horner speaks directly to Google's financial incentives to be complicit:

> Right now I make like $10,000 a month from AdSense. I know ways of getting hooked up under different names and sites. So probably if they cracked down, I would try different things. I have at least 10 sites right now. If they crack down on a couple, I'll just use others. […] Plus, Facebook and AdSense make a lot of money from [advertising on fake news sites] for them to just get rid of it. They'd lose a lot of money. (qtd. in Dewey)

Google and Facebook both released statements in mid-November 2016 that they would fight against fake news (Love and Cooke) by restricting their advertising, with Google claiming the ban would come into effect "imminently" (Wingfield et al.). However, a number of different reports, including *The Wall Street Journal* and *Media Matters for America*, detailed "advertisements placed by Google on at least 24 websites that have a track record for pushing fake news stories – stories with fabricated information packaged to appear as a legitimate news story" (Suen et al.). The *Media Matters for America* report was conducted on 12 December, almost a full month after Google stated it would ban fake news sites from their revenue stream on 14 November.

The most popular fake news during the 2016 election favoured Trump over Hillary Clinton. Allcott and Gentzkow collected a database of fake news stories from three major repositories and found that their database contained "115 pro-Trump

fake stories that were shared on Facebook a total of 30 million times, and 41 pro-Clinton fake stories shared a total of 7.6 million times" (212). However, this number does not necessarily mean that all those who were sharing it were Trump supporters. Because so many of the fake news creators aimed to create stories with the largest potential to be widely shared, rather than politically persuade, it may be the case that the most viral stories were those that mobilised online users from across the political spectrum to share the article through outrage, disbelief, agreement or any other motivating factor. Using "previously reported statistics for the ratio of page visits to shares of stories on social media" Allcott and Gentzkow calculate that the combined "38 million shares of fake news in [their] database translates into 760 million instances of a user clicking through and reading a fake news story, or about three stories read per American adult" (212). In addition, this figure does not even include users who partially read stories or view headlines within Facebook, Twitter, or Google News directly. These numbers indicate that fake news is not a small issue or one that is easily dismissed. These figures also emphasise the extensive reach that these fake news article had, the accountability of which, also lies with Google and Facebook. Not only did these two companies profit from the surfeit of fake news, but actively promoted it, whether algorithmically or intentionally, through search engine results and social media feeds.

The wider changes in the online access to journalism mean that traditional media outlets now compete for revenue on the same terms as fake news sites. Without the institutional stability of print sales[16] or other regular revenue support, traditional media organisations are drawn into writing in ways that attract attention and generate as many individual page views as possible, in order to increase

[16] A Pew Research study found that between 2005 and 2015 US newspaper "weekday circulation has fallen 17% and ad revenue more than 50%" (Mitchell and Matsa).

advertising revenue. Combined with the media-specific characteristics of search engine results and social media feeds that decontextualise individual articles and present a diverse range of content, Google and Facebook encourage a logic that incentives clickbait headlines. There is a clear difference between misleading articles and fake news; however, in many ways fake news is a gross exaggeration of current online news practices rather than an activity that occupies a completely separate domain. As long as profits are tied directly to how much an article is shared or viewed then very particular kinds of media content will continue to be incentivised over others.

Fake news is just one example to consider when investigating how Google creates avenues for profit and how Google's economics co-depend on other online institutions, in particular, Facebook. These mutual incentives promote particular kinds of relationships between individual web users and online institutions such as Facebook and Google in a way that promotes Post-Fordist interactions. Immaterial labour is becoming a more expansive domain. As Maurizio Lazzarato argues,

> If production today is directly the production of a social relation, then the "raw material" of immaterial labor is subjectivity and the "ideological" environment in which this subjectivity lives and reproduces. The production of subjectivity ceases to be only an instrument of social control (for the reproduction of mercantile relationships) and becomes directly productive, [immaterial workers] satisfy a demand by the consumer and at the same time establish that demand. (143)

The interrelations of Google, Facebook, creators of fake news, and users

demonstrate how economic value is being created and harvested from relationships that many online users would not see as primarily economic. These examples also demonstrate how the rise of immaterial labour by no means replaces traditional material labour, but that increasingly the spare time of individuals is spend working, in a Post-Fordist sense, to increase the profits of particular institutions and creators. Even if a user shares a piece of fake news out of outrage, they have still acted within a system that uses advertising revenue to financially compensate its creator, the social media network, and the company that supported its advertising. It is important that web users understand that the web is structured around financial incentives and that, collectively, the actions of following links and sharing pages are intrinsically economic and carry significant consequences for the future of the global information ecology.

Conclusion

This chapter demonstrates how Google's models of advertising are central to its search engine and, more broadly, its structuring of the web. The implications of many of the issues outlined in this chapter are significant; many of the consequences may seem inescapable or at least difficult to influence. For example, the future of diminishing language diversity appears inevitable, if current economic incentives stay as they are. Nevertheless, this chapter highlights that there are still potential routes for change. The preceding argument has articulated why Post-Fordist theory provides a useful context for understanding Google's relationship to contemporary digital capitalism; grounding an understanding through terminology such as immaterial labour and the general intellect can recontextualise current practices and clarify certain intersections of information, capital, and individual agency. When considered alongside Williams' history of advertising, this approach clarifies the current state of capitalism and the particular influence that Google possesses. Describing Google as an advertising company, while appreciating the historically situated nature of this form of advertising, reframes the motives behind Alphabet's other projects in terms of data collection and highlights how that company is shaping much more than search engine results. Chapter two placed Google's project within a broader historical narrative, while chapter one highlighted the technical differences and similarities between traditional questions and querying the world through web searches. An economic understanding does not overturn those narratives; however, Google's influence as a globally powerful private business sets it apart from any of its historical antecedents.

Finally, the example of fake news and its expansion during the 2016 US

presidential election, although alarming, should serve as an illustration of how online economic incentives produce significant effects that implicate all individuals on a global scale. In summary, this chapter demonstrates that contemporary digital capitalism is not divorced from the non-digital world, but rather represents a significant part of a larger shift within twenty-first-century capitalism. Google's influence on this shift is considerable, and therefore, proper Internet citizenship requires an understanding of the sway that this particular company holds over current and future digital practices. This understanding is essential if we are to build and sustain a web that reflects the diverse cultures of its global users, rather than the economic incentives of a single company.

Conclusion

This thesis has argued that search engines embody and reproduce particular social and cultural values, from their most basic underlying code to their most complex methods of aggregating data. Google's search engine has shaped the content of the web and the ways that individuals think, remember, and act online, on a global scale. Understanding search engines requires a time-specific and technologically robust grasp of contemporary digital culture; however, placing the issues that develop from studying Google into an historical and philosophical context demonstrates that many of our present-day concerns rearticulate long-held debates. Finally, Google's practices demonstrate how global digital culture is being directed by a singular economic narrative, which may have dramatic consequences for the social, cultural, and linguistic diversity that can survive online. This conclusion reflects on the original research questions of the project and provides an interconnected set of speculative concerns regarding the future of Google's dominance, search engine technology and the web more generally.

1.0 Research Questions Addressed

The original research questions, outlined in the introduction, have been addressed in the following way:

How do individuals interrogate and navigate the world around them?

Chapter one demonstrated that the technical structure underpinning search engine queries provides a new mode of linguistic enquiry, which has influenced the overall content of the web. Drawing information retrieval studies into contact with philosophy demonstrates that while queries have a particular set of technological affordances they are also part of a broader tradition of asking questions, which carry an existing set of analytical issues.

How do technologies and social institutions facilitate how we think and remember?

Chapter two showed that memory has always been embedded within and facilitated by technosocial practices and institutions. Placing the findings of contemporary psychological research regarding transactive memory within an historical context counteracts any conclusion that describes the mnemonic relationship between individuals and Google as unprecedented. However, given the conclusions of chapter five, concerning the economic structure of search engines, this history serves to highlight the particularities of the privatised nature of Google's position.

How culturally situated is knowledge; are there epistemological truths that transcend social environments?

Chapter three provided an insight into the way in which Google's Autocomplete tool situates users within an algorithmically determined context. This chapter also

revealed the importance of Google's criterion of *relevance* as an underlying logic in which epistemological landscapes are established by large-scale aggregations of data. Google's attitude towards the question of objective truth appears to be their belief that more data always leads to more *relevant* results. However, as the examples demonstrating the perpetuation of stereotypes show, such a metric carries complex cultural questions that cannot simply be answered by increasing quantities of data.

How does technological expansion fit within wider questions of globalisation?
Chapter four demonstrated that various contextual factors affect Google's results. That in 2015, Google's results significantly differed between languages, locations, and the linguistic patterns of individuals, emphasised the centrality of *relevance*. The repeated study of 2017 demonstrated that the factors used to provide results constantly shift and that wider questions of globalisation and a shared epistemological landscape are being shaped by hidden and automated criteria.

How do technological discourses shape the global flows of information and capital?
Chapter five showed that the economic basis of ubiquitous technology can create a range of effects, many of which may be unintentional and some particularly undesirable. Post-Fordist theory provided a framework to describe a new relationship between capital and information that Google reinforces and benefits from. The example of fake news demonstrates that technological structures create social conditions that can have wide-ranging implications for digital culture and a wider global landscape.

2.0 Google's Dominance and the Role of Search Engines

Apart from some national-specific search engines, for example Baidu in China and Yandex in Russia, Google dominates the search engine landscape. This is not only the case in terms of market share but also in terms of representing a gold standard in which competitors try to replicate, rather than offer an alternative; this was addressed in chapter four by highlighting evidence that in 2011 Bing had illegally copied Google's results wholesale. However, making a claim regarding the general outlook that Google Search presents is difficult. In some ways, Google Search is becoming more personalised through their emphasis on relevance. However, chapter four's 2017 repeat of the 2015 study suggested that search results were becoming more homogenised, although a much larger study would be needed in order to argue that this trend was widespread. If Google Search results are becoming more standardised it could be because of Google's move to embrace of neural network machine learning. It could also be a pre-emptive way of shoring up their territory of 'objective' web access in contrast to the personalised content curation of social media providers, such as Twitter and Facebook. In an interview in 2012, Mark Zuckerberg explicitly spoke about using Facebook's existing user data to build a search engine to rival Google: "Facebook is pretty uniquely positioned to answer the questions people have. At some point we'll do it. We have a team working on it" (quoted in Olanoff). Such a search engine may replicate the hyper-personalisation of Facebook's current approach; Google's development of Google Now and 'the feed', discussed in chapter three, may also represent a movement in this direction.

Putting speculation regarding specific companies to one side, one of the central questions of this thesis will continue to be important: when using technology to ask questions, should the answers provided be contextually relevant and culturally contingent, or should some types of discourses dominate on a global scale? The historical narrative of chapter two and the philosophical positions of Plato and Aristotle will continue to be relevant to current debates regarding ubiquitous technology. A particularly problematic part of digital culture today is the echo chamber nature of some online discourse whether it be the rise of the alt-right, conspiracy theory, hate groups, or campaigns of targeted harassment such as the #gamergate movement of 2014.[1] These are complex social phenomena but understanding the underlying logic of search engines clearly describes one piece of that puzzle, which is that the kind of language individuals use in their search queries determines the kinds of answers they receive, the institutions that have authority, and the kinds of community spaces or forums presented to them, which might reinforce existing attitudes. The conclusions of chapter five, regarding the political effects of Google's economic incentives, describe a situation in which a particular technological structure has facilitated the rise of certain types of content. It would be an overstatement to claim that search engines *caused* the polarised social trends of 2016; however, analysing the rise of fake news demonstrates one of the ways in which search engine technology is enmeshed within a wider social, political, and economic sphere.

[1] See Angela Nagle's *Kill All Normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*, George Hawley's *Making Sense of the Alt-Right*, and David *Neiwert's Alt America: The Rise of the Radical Right in the Age of Trump*.

2.1 The Future of the Web

Google's search engine has become the central way of accessing the web, but as online activities move towards the use of mobiles and tablets, many have argued that the importance of the web may diminish and instead give way to a completely app-based digital environment. Apps provide walled gardens for certain companies to more directly control the kinds of content users receive, while also preventing individuals from using software that blocks advertising online. In 2014, Marcus Wohlsen argued that if

> app-happy mobile devices become the primary way we compute, the good old browser becomes irrelevant. The hyperlinked, free-flowing, egalitarian, and ubiquitous world wide web will fade away. Instead, digital existence will mostly transpire within the more self-contained domains of individual apps, which offer their creators the flexibility and power of building right into the mobile operating systems. We will still have the internet, but it won't be the same wherever you use it. And some will have more power over it than others.

As has been demonstrated throughout this thesis, the web is far from egalitarian or even globally consistent. Chapter one outlined how Google reinforces existing hierarchies and chapter four demonstrated that the landscape of the web does already shift between contexts. However, if the online environment did become primarily app-based these economic and power disparities would become more dramatic. Many online institutions, for example Facebook and Google, are trying hard to push individuals to use their real-life credentials and to enable these profiles

to function as proxies for other sites (when creating an account on an unrelated site the message, “would you like to sign up with Facebook?” has become commonplace). This current trajectory has the potential to shape the nature of enquiry in the future; if users are always identified when online, the potential for personalising results increases. If apps did replace the web, Google would still have major control of the online environment, due to their ownership of Android’s app store, which represents a monopoly on that operating system, an operating system with a significantly high global market share (see fig. 1).

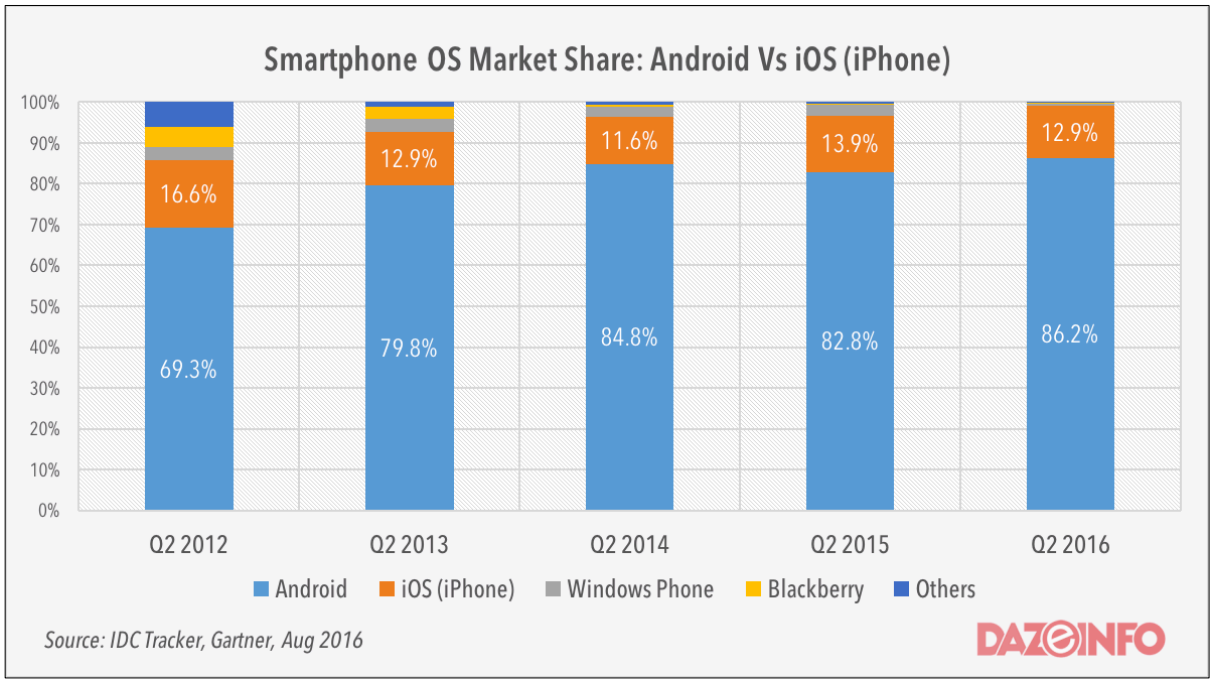


Figure 1. Android’s share of the Operating System (OS) market is increasing and in 2016 represented 86.2% globally. Source: “Smartphone OS Market Share, 2012-2016” (see Chau et al.). Accessed 11/10/2016.

An app-based environment also means that the control of information becomes increasingly tied to hardware. Which institutions control these informational flows becomes an increasingly economic and legal question. In addition, the media

channels available for small-scale content creators become less feasible and the discovery of such work becomes more difficult, other than through partnerships with larger or more established institutions. A central function of search engines is discovery and although content creators already rely on Google's hegemony, the online landscape has the potential to become even more centralised.

2.2 Ubiquity of Data

Google and other companies, such as Apple, Microsoft and Amazon, are investing in ways to capture increasing amounts of data. Throughout this thesis, I have highlighted the centrality of Google's mission statement: "to organize the world's information and make it universally accessible and useful". Web search may be the current economic engine of Alphabet but it is only one part of a wider narrative of data collection – including their other projects such as Google Books, Maps, Street View and others – the logic of which conceptualises the world in terms of data to be collected and put to use. Ubiquitous data collection is becoming normalised. An environment in which the default is for information to be automatically recorded rather than forgotten has particular implications for chapters two and three. In particular, an increase in data-collection may dramatically change how individuals and communities conceptualise memory, as well the wider expectations regarding how such information should be used. Future research of digital culture will need to foreground these wider debates regarding how much automated aggregation of information should dictate the wider technological landscape. This debate will occur in a number of domains, however, the final chapter of this thesis highlights that these negotiations, first and foremost, will be structured through dominant economic

relationships. A central implication of this research is that, although many of the social and philosophical questions regarding search engines have been central throughout history, digital culture is shifting towards a particular environment that benefits particular companies over the increasing number of online users.

Bibliography

Alberth, William P., Jr. *Coupling an Electronic Skin Tattoo to a Mobile Communication Device*. WO2013166377 A1, 7 Nov. 2013, www.google.co.uk/patents/WO2013166377A1. Patent.

Ali, Rashid, and M. M. Sufyan Beg. “Automatic Performance Evaluation of Web Search Systems Using Rough Set Based Rank Aggregation.” *Proceedings of the First International Conference on Intelligent Human Computer Interaction*, Springer, 2009, pp. 344–58, doi:10.1007/978-81-8489-203-1_34.

Allcott, Hunt, and Matthew Gentzkow. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives*, vol. 31, no. 2, May 2017, pp. 211–36, doi:10.1257/jep.31.2.211.

Allison, Lloyd. “Tries.” *Lloyd Allison Computing*, www.allisons.org/ll/AlgDS/Tree/Trie/. Accessed 13 Aug. 2017.

Alphabet. “Alphabet Investor Relations: 2016 Annual Report.” *Alphabet Investor Relations*, 2017, abc.xyz/.

Ambrosino, Brandon. “I Am Gay – but I Wasn’t Born This Way.” *BBC Future*, 28 June 2016, www.bbc.com/future/story/20160627-i-am-gay-but-i-wasnt-born-this-way.

Amin, Ash, editor. *Post Fordism: A Reader*. John Wiley & Sons, 1994.

Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. 1983. Verso, 1991.

Anderson, Chris. “The Long Tail: About Me.” *The Long Tail*, www.longtail.com/about.html. Accessed 11 June 2017.

"AOL 2006 Query Log." *McGill's Centre for Intelligent Machines*, www.cim.mcgill.ca/~dudek/206/Logs/AOL-user-ct-collection/. Accessed 29 Oct. 2017.

Aquinas, Thomas. "Quaestiones Disputatae De Veritate, Q. 2-4." *Corpus Thomisticum*, www.corpusthomisticum.org/qdv02.html. Accessed 28 Oct. 2017.

Assange, Julian. *When Google Met WikiLeaks*. OR Books, 2016.

Auchard, Eric. "Trademark Plaintiff Drops Suit Vs. Google Over Ads." *Reuters*, 4 Sept. 2007, www.reuters.com/article/us-google-trademark-idUSN0336124420070904.

Auletta, Ken. *Googled: The End of the World as We Know It*. Penguin Books, 2009.

Azuma, Hiroki. *General Will 2.0: Rousseau, Freud, Google*. Translated by John Person, Vertical, 2014.

Baker, Paul, and Amanda Potts. "'Why Do White People Have Thin Lips?' Google and the Perpetuation of Stereotypes Via Auto-Complete Search Forms." *Critical Discourse Studies*, vol. 10, no. 2, May 2013, pp. 187–204, doi:10.1080/17405904.2012.744320.

Baker, Steven. "Helping Computers Understand Language." *Official Google Blog*, 19 Jan. 2010, googleblog.blogspot.com/2010/01/helping-computers-understand-language.html.

Baraniuk, Chris. "Brexit: 'Surge' in Searches on Irish Passports, Says Google." *BBC News*, 24 June 2016, www.bbc.co.uk/news/technology-36619223.

Bar-Ilan, Judit. "On the Overlap, the Precision and Estimated Recall of Search Engines: A Case Study of the Query 'erdos.'" *Scientometrics*, vol. 42, no. 2, June 1998, pp. 207–28, doi:10.1007/BF02458356.

Battelle, John. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Nicholas Brealey Publishing, 2006.

Becker, Konrad, and Felix Stalder, editors. *Deep Search: The Politics of Search Beyond Google*. Studien Verlag, 2009.

Belew, Richard K. *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*. Cambridge UP, 2001.

Berardi, Franco "bifo." *And: Phenomenology of the End*. MIT Press, 2015.

Beres, Damon. "After Brexit, U.K. Residents Google 'What Is The EU?'" *The Huffington Post*, 24 June 2016, www.huffingtonpost.com/entry/brexit-what-is-the-eu-google_us_576d2dfee4b0dbb1bbba3911.

Berry, Michael W., and Murray Browne. *Understanding Search Engines: Mathematical Modeling and Text Retrieval*. Society for Industrial and Applied Mathematics, 2005.

Blumenthal, Eli. "Google Search Spike Casts Doubts on U.K. Brexit Knowledge." *USA Today*, 24 June 2016, www.usatoday.com/story/tech/2016/06/24/google-search-spike-casts-doubts-uk-brexit-knowledge/86336834/.

Borges, Jorge Luis. "On Exactitude in Science." *Collected Fictions*, translated by Andrew Hurley, Penguin Books, 1999, pp. 325–26.

---. "The Total Library." *The Total Library: Non-Fiction 1922-1986*, translated by Esther Allen and Suzanne Levine, Penguin Books, 2001, pp. 214–16.

Boutang, Prof Yann Moulier. *Cognitive Capitalism*. Polity, 2012.

Brandt, Richard L. *The Google Guys: Inside the Brilliant Minds of Google Founders Larry Page and Sergey Brin*. Portfolio, 2011.

Brier, Soren. *Cybersemiotics: Why Information Is Not Enough*. University of Toronto Press, 2014.

Brin, Sergey, and Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Seventh International World-Wide Web Conference* 1998, Stanford InfoLab Publication Server, ilpubs.stanford.edu:8090/361/.

Broder, Andrei. “A Taxonomy of Web Search.” *ACM SIGIR Forum*, vol. 36, Association for Computing Machinery, 2002, pp. 3–10, doi:10.1145/792550.792552.

Burke, Peter. *A Social History of Knowledge: From Gutenberg to Diderot*. Polity, 2000.

Buscher, Georg, et al. “The Good, the Bad, and the Random: An Eye-Tracking Study of Ad Quality in Web Search.” *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, 2010, pp. 42–49, doi:10.1145/1835449.1835459.

Bute, M. S., et al. “Evaluating Search Effectiveness of Some Selected Search Engines.” *Global Journal of Pure and Applied Sciences*, vol. 23, 2017, pp. 139–49, www.globaljournalseries.com/index.php/gjpas/article/view/330.

Carr, Nicholas. “Is Google Making Us Stupid?” *The Atlantic*, July/August Issue, 2008, www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/.

---. *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember*. Atlantic Books, 2011.

Carruthers, Mary. “How to Make a Composition: Memory-Craft in Antiquity and in the Middle Ages.” *Memory: Histories, Theories, Debates*, edited by Susannah Radstone and Bill Schwarz, Fordham UP, 2010, pp. 15–29.

---. *The Book of Memory: A Study of Memory In Medieval Culture*. 2nd ed., Cambridge UP, 2008.

Chan, Rosalie. “The U.K. Is Googling What the E.U. Is Hours After It Voted to Leave.” *Time*, 24 June 2016, time.com/4381612/uk-brexit-google-what-is-the-eu/.

Chau, Melissa, et al. “Smartphone OS Market Share, 2012-2016.” *International Data Corporation*, www.idc.com/promo/smartphone-market-share/os. Accessed 6 Dec. 2017.

Chitika Insights. “The Value of Google Result Positioning.” *Chitika Online Advertising Network*, 7 June 2013, chitika.com/google-positioning-value.

Chopra, Shruti Tripathi. “Google Hit with Record €2.4bn Fine by the European Commission for Promoting Its Own Shopping Comparison Service at the Top of Search Results.” *City A.M.*, 27 June 2017, www.cityam.com/267401/google-hit-record-eur24bn-fine-european-commission.

Chu, Heting, and Marilyn Rosenthal. “Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology.” *Proceedings of the ASIST Annual Meeting*, vol. 33, 1996, pp. 127–35, www.learntechlib.org/p/83947.

Clark, Andy, and David J. Chalmers. “The Extended Mind.” *Analysis*, vol. 58, no. 1, 1998, pp. 7–19.

Clarke, Sarah, and Peter Willett. “Estimating the Recall Performance of Web Search Engines.” *Aslib Proceedings*, vol. 49, no. 7, July 1997, pp. 184–89, doi:10.1108/eb051463.

Cohen, Dan. “What the Google Books Victory Means for Readers.” *The Atlantic*, Oct. 2015, www.theatlantic.com/technology/archive/2015/10/what-the-google-books-victory-means-for-readers-and-libraries/411910/.

Cole, Jeffery I., et al. *The UCLA Internet Report, Surveying the Digital Future*. UCLA Center for Communication Policy, 2003.

Coleman, Janet. *Ancient and Medieval Memories: Studies in the Reconstruction of the Past*. Cambridge UP, 1992.

Couliano, Ioan P. *Eros and Magic in the Renaissance*. 2nd ed., University of Chicago Press, 1987.

Creech, Jimmy. “What Does the Bible Say About Homosexuality?” *Human Rights Campaign*, www.hrc.org/resources/what-does-the-bible-say-about-homosexuality/. Accessed 5 Dec. 2017.

Cutts, Matt. “How Search Works.” *YouTube*, uploaded by Google, 4 March 2010, www.youtube.com/watch?v=BNHR6IQJGZs.

Dean, Brain. “Google’s 200 Ranking Factors: The Complete List.” *Backlinko*, 5 Nov. 2016, backlinko.com/google-ranking-factors.

Débarre, Florence. “Are Google Suggestions Sexist?” *GitHub*, 7 June 2016, htmlpreview.github.io/?raw.githubusercontent.com/flodebarre/GoogleSuggestions/master/report.html.

Derrida, Jacques. “Typewriter Ribbon: Limited Ink (2).” *Without Alibi*, translated by Peggy Kamuf, Stanford UP, 2002, pp. 71–160.

Dewey, Caitlin. “Facebook Fake-News Writer: ‘I Think Donald Trump Is in the White House because of Me.’” *The Bangor Daily News*, 17 Nov. 2016, bangordailynews.com/2016/11/17/news/nation/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/.

Diakopoulos, Nicholas. “Rage Against the Algorithms.” *The Atlantic*, Oct. 2013, www.theatlantic.com/technology/archive/2013/10/rage-against-the-algorithms/280255/.

Diaz, Alejandro. “Through the Google Goggles: Sociopolitical Bias in Search Engine Design.” *Web Search: Multidisciplinary Perspectives*, Springer, 2008, pp. 11–34, doi:10.1007/978-3-540-75829-7_2.

Ding, Wei, and Gary Marchionini. “A Comparative Study of Web Search Service Performance.” *Proceedings of the ASIST Annual Meeting*, vol. 33, 1996, pp. 136–40.

Dischler, Jerry. “Building for the Next Moment.” *Inside AdWords*, 5 May 2015, adwords.googleblog.com/2015/05/building-for-next-moment.html.

Dong, Xin, et al. “Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion.” *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2014, pp. 601–610, dl.acm.org/citation.cfm?id=2623623.

Draaisma, Douwe. *Metaphors of Memory: A History of Ideas about the Mind*. Translated by Paul Vincent, Cambridge UP, 2000.

Duguid, Paul. “Search before Grep. A Progress from Closed to Open?” *Deep Search: The Politics of Search Beyond Google*, edited by Konrad Becker and Felix Stalder, Studien Verlag, 2009, pp. 13–31.

Dyson, George. *Turing’s Cathedral: The Origins of the Digital Universe*. Penguin Books, 2012.

Edwards, Douglas. *I’m Feeling Lucky: The Confessions of Google Employee Number 59*. Mariner Books, 2012.

Effrat, Jonathan, et al. “Faster Is Better on Google Suggest.” *Official Google Blog*, 20 May 2009, googleblog.blogspot.com/2009/05/faster-is-better-on-google-suggest.html.

Elfrink, Greg. “The Adsense Business Model Explained.” *Empire Flippers*, 11 May 2016, empireflippers.com/adsense-business-model-explained/.

European Commission. “State Aid: Commission Endorses €120 Million Aid for German R&D Project THESEUS.” *European Commission Press Release Database*, 19 July 2007, europa.eu/rapid/press-release_IP-07-1136_en.htm.

Fallows, Deborah. “Internet Searchers Are Confident, Satisfied and Trusting – But They Are Also Unaware and Naïve.” *Pew Internet & American Life Project*, 2005.

Felten, Ed. “Googlocracy in Action.” *Freedom to Tinker*, 3 Feb. 2004, freedom-to-tinker.com/2004/02/03/googlocracy-action/.

Fisher, Mark. *Capitalist Realism: Is There No Alternative?* Zero Books, 2009.

Fortune. “Fortune 500 Companies 2017: Who Made the List.” *Fortune*, 11 June 2017, fortune.com/fortune500/list/.

Fox, Susannah. “Search Engines.” *Pew Research Center: Internet, Science & Tech*, 3 July 2002, www.pewinternet.org/2002/07/03/search-engines/.

Franklin, Curt. “How Internet Search Engines Work.” *How Stuff Works*, 27 Sept. 2000, computer.howstuffworks.com/internet/basics/search-engine.htm.

Fry, Jenny, et al. “Search Engines and Expertise About Global Issues: Well-Defined Landscape or Undomesticated Wilderness?” *Web Search: Multidisciplinary Perspectives*, Springer, 2008, pp. 255–75, doi:10.1007/978-3-540-75829-7_14.

Fung, Brian. “The British Are Frantically Googling What the E.U. Is, Hours after Voting to Leave It.” *Washington Post*, 24 June 2016, www.washingtonpost.com/news/the-switch/wp/2016/06/24/the-british-are-frantically-googling-what-the-eu-is-hours-after-voting-to-leave-it/.

Gadamer, Hans-Georg. *Truth and Method*. Translated by Joel Weinsheimer and Donald G. Marshall, Bloomsbury, 2013.

Gannes, Liz. “Nearly a Decade Later, the Autocomplete Origin Story: Kevin Gibbs and Google Suggest.” *AllThingsD*, 23 Aug. 2013, allthingsd.com/20130823/nearly-a-decade-later-the-autocomplete-origin-story-kevin-gibbs-and-google-suggest/.

Garrahan, Matthew. “Advertising: Facebook and Google Build a Duopoly.” *Financial Times*, 23 June 2016, www.ft.com/content/6c6b74a4-3920-11e6-9a05-82a9b15a8ee7.

Gatti, Hilary. *Giordano Bruno and Renaissance Science: Broken Lives and Organizational Power*. Cornell UP, 2002.

Geertz, Clifford. “Thick Description: Toward an Interpretive Theory of Culture.” *The Interpretation of Cultures*, Basic Books, 1977, pp. 3–32.

Gerhart, Susan. “Do Web Search Engines Suppress Controversy?” *First Monday*, vol. 9, no. 1, Jan. 2004, firstmonday.org/ojs/index.php/fm/article/view/1111.

Germann, Ulrich, et al. “Tightly Packed Tries: How to Fit Large Models into Memory, and Make Them Load Fast, Too.” *Proceedings of the Workshop on Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, Association for Computational Linguistics, 2009, pp. 31–39, dl.acm.org/citation.cfm?id=1621952.

Gibbs, Kevin. “I’ve Got a Suggestion.” *Official Google Blog*, 10 Dec. 2004, googleblog.blogspot.com/2004/12/ive-got-suggestion.html.

Globalization Group Inc. “Top Languages by GDP.” *The International Business Edge*, 2010, www.globalization-group.com/edge/2010/03/top-languages-by-gdp/.

Godfrey, Alex. “Michael Cera: From Superbad Star to Mescaline-Soaked Drug Adventurer.” *The Guardian*, 10 Jan. 2014, www.theguardian.com/film/2014/jan/10/michael-cera-crystal-fairy.

Goel, Shikha, and Sunita Yadav. “An Overview of Search Engine Evaluation Strategies.” *International Journal of Applied Information Systems*, vol. 1, no. 4, Feb. 2012, pp. 7–10, doi:10.5120/ijais12-450156.

Goldman, Eric. “Search Engine Bias and the Demise of Search Engine Utopianism.” *Yale Journal of Law and Technology*, vol. 8, no. 1, Jan. 2006, digitalcommons.law.yale.edu/yjolt/vol8/iss1/6.

Goldstein, Rebecca. *Plato at the Googleplex: Why Philosophy Won't Go Away*. Vintage, 2015.

Google. "Advertiser Guide: Working with a Third-Parties." *Google AdWords*, www.google.com/intl/en_uk/adwords/thirdpartypartners/. Accessed 8 June 2017.

---. "AdWords Policies." *Google Advertising Policies Help*, support.google.com/adwordspolicy/answer/6008942?hl=en-GB. Accessed 8 June 2017.

---. "Deep Dream: About." *Deep Dream Generator*, deepdreamgenerator.com/. Accessed 14 Aug. 2017.

---. "Gambling and Games." *Google Advertising Policies Help*, support.google.com/adwordspolicy/answer/6018017?hl=en-GB. Accessed 9 June 2017.

---. "Google - Year In Search 2016." *YouTube*, uploaded by Google, 14 Dec. 2016, www.youtube.com/watch?v=KIViy7L_lo8.

---. "Google: About Us." *Google*, www.google.com/about/. Accessed 8 June 2015.

---. "Google Acquires Technology Assets of Outride Inc. – News Announcements." *Official Google Blog*, 20 Sept. 2001, googlepress.blogspot.co.uk/2001/09/google-acquires-technology-assets-of.html.

---. "Google Inside Search: Autocomplete." *Google Inside Search*, 18 Apr. 2013, web.archive.org/web/20130418015232/support.google.com/websearch/answer/106230?hl=en.

---. "Google Search Help: Autocomplete." *Google Search Help*, 9 Apr. 2015, web.archive.org/web/20150409060419/support.google.com/websearch/answer/106230?hl=en.

---. “Google Search Help: Search Using Autocomplete.” *Google Search Help*, support.google.com/websearch/answer/106230. Accessed 13 Aug. 2017.

---. “Google Search Results Page.” *Google Search Help*, support.google.com/websearch/answer/35891?hl=en. Accessed 28 Oct. 2017.

---. “Google: Technology.” *Google*, www.google.com/technology. Accessed 20 Oct. 2017.

---. “Google’s Year in Search 2016.” *Google Trends*, g.co/trends/HUzzl. Accessed 28 Oct. 2017.

---. “Our New Search Index: Caffeine.” *Official Google Blog*, 8 June 2010, googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html.

---. “Project Loon.” *X Company*, x.company/loon/. Accessed 9 June 2017.

---. “Transparency Report, Germany.” *Google Transparency Report*, transparencyreport.google.com/government-removals/by-country/DE. Accessed 27 Aug. 2017.

---. “Welcome to the 2017 Google Online Marketing Challenge.” *Google*, www.google.com/onlinechallenge/. Accessed 8 June 2017.

@GoogleTrends. “+250% Spike In ‘what Happens If We Leave the EU’ in the Past Hour Google.com/Trends/Explore...” *Twitter*, 23 June 2016, twitter.com/googletrends/status/746137920940056578?lang=en.

---. “‘What Is the EU?’ is the Second Top UK Question on the EU since the #EURefResults Were Officially Announced.” *Twitter*, 24 June 2016, twitter.com/googletrends/status/746303118820937728?lang=en.

Graves, Allison. “Did Google Adjust Its Autocomplete Algorithm to Hide Hillary Clinton’s Problems?” *Politifact*, 23 June 2016, www.politifact.com/punditfact/statements/2016/jun/23/andrew-napolitano/did-google-adjust-its-autocomplete-algorithm-hide-/.

Greenberg, Andy. “Hacker Lexicon: What Is the Dark Web?” *Wired*, 19 Nov. 2014, www.wired.com/2014/11/hacker-lexicon-whats-dark-web/.

Griffin, Andrew. “What Happens If We Leave the EU? Google Searches Surge as People Realise They Don’t Know What Brexit Actually Means.” *The Independent*, 24 June 2016, www.independent.co.uk/life-style/gadgets-and-tech/news/what-happens-if-we-leave-the-eu-google-searches-surge-as-people-realise-they-dont-know-what-brexit-a7100596.html.

Grondin, Jean. *The Philosophy of Gadamer*. Translated by Kathryn Plant, Routledge, 2014.

Groys, Boris. “Google: Words Beyond Grammar.” *In the Flow*, Verso Books, 2016, pp. 147–56.

Halavais, Alexander. *Search Engine Society*. Polity, 2008.

Hall, Jonathan V., and Alan B. Krueger. “An Analysis of the Labor Market for Uber’s Driver-Partners in the United States.” *Working Papers (Princeton University Industrial Relations Section)*, no. 587, Jan. 2015, arks.princeton.edu/ark:/88435/dsp010z708z67d.

Hardt, Michael, and Antonio Negri. *Empire*. Harvard UP, 2001.

Hargittai, Eszter. “A Framework for Studying Differences in People’s Digital Media Uses.” *Grenzenlose Cyberwelt?*, 2007, pp. 121–136.

---. “Beyond Logs and Surveys: In-Depth Measures of People’s Web Use Skills.” *Journal of the American Society for Information Science and Technology*, vol. 53, no. 14, Dec. 2002, pp. 1239–44, doi:10.1002/asi.10166.

---. “The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction.” *Journal of Computer-Mediated Communication*, edited by Eszter

Hargittai, vol. 12, no. 3, Apr. 2007, pp. 769–77, doi:10.1111/j.1083-6101.2007.00349.x.

Hariri, Nadjla, et al. “The Precision and Recall of General Search Engines in Retreival of Images Related to Endocrine Diseases.” *Iranian Journal of Endocrinology and Metabolism*, vol. 17, no. 2, Aug. 2015, pp. 97–104.

Harman, Graham. *Heidegger Explained: From Phenomenon to Thing*. Open Court, 2007.

Harris, Julian. “EU Watchdog Encourages Firms to Sue Google for Anti-Competitive Practices.” *City A.M.*, 2 July 2017, www.cityam.com/267672/european-commission-antitrust-watchdog-margrethe-vestager.

Havelock, Eric. *Preface to Plato*. Belknap Press, 1963.

Havens, John C. “The Connected Home May Become the Collected Home.” *Slate*, July 2014, www.slate.com/blogs/future_tense/2014/07/31/nest_google_acquisition_the_connected_home_may_be_the_collected_home.html.

Hawley, George. *Making Sense of the Alt-Right*. Columbia UP, 2017.

Heath, Taliver Brooks. *Advertising Based on Environmental Conditions*. US8138930 B1, 20 Mar. 2012, www.google.co.uk/patents/US8138930. Patent.

Heidegger, Martin. “Letter on ‘Humanism.’” *Pathmarks*, edited by William McNeil, translated by Frank A. Capuzzi, Cambridge UP, 1998, pp. 239–76.

Herrman, John. “Media Websites Battle Faltering Ad Revenue and Traffic.” *The New York Times*, 17 Apr. 2016, www.nytimes.com/2016/04/18/business/media-websites-battle-falteringad-revenue-and-traffic.html.

Hess, Amanda. “‘Google Autocomplete Is’: Not as Fun as It Used to Be.” *Slate*, 18 Mar. 2015,

www.slate.com/articles/technology/users/2015/03/google_autocomplete_not_as_weird_dark_or_fun_as_it_used_to_be.html.

Heyman, Stephen. “Google Books: A Complex and Controversial Experiment.” *The New York Times*, 28 Oct. 2015, www.nytimes.com/2015/10/29/arts/international/google-books-a-complex-and-controversial-experiment.html.

Hillis, Ken, et al. *Google and the Culture of Search*. Routledge, 2012.

Hölscher, Christoph, and Gerhard Strube. “Web Search Behavior of Internet Experts and Newbies.” *Proceedings of the 9th International World Wide Web Conference on Computer Networks: The International Journal of Computer and Telecommunications Networking*, vol. 33, 2000, pp. 337–346, doi:10.1016/S1389-1286(00)00031-1.

“Homosexuality.” *Wikipedia*, en.wikipedia.org/wiki/Homosexuality. Accessed 28 Nov. 2017.

Hook, Justin. *Google Feud*. www.googlefeud.com/. Accessed 1 Dec. 2017.

Horner, Paul (pseud. Jimmy Rustling). “Donald Trump Protester Speaks Out: ‘I Was Paid $3,500 To Protest Trump’s Rally.’” *ABC News*, 22 Nov. 2016, abcnews.com.co/donald-trump-protester-speaks-out-i-was-paid-to-protest/.

Hotchkiss, Gord, et al. “Eye Tracking Study.” *Enquiro Search Solutions Inc*, June 2005, searchengineland.com/figz/wp-content/seloads/2007/09/hotchkiss-eye-tracking-2005.pdf.

Huebner, Bryce. “Transactive Memory Reconstructed: Rethinking Wegner’s Research Program.” *Southern Journal of Philosophy*, vol. 54, no. 1, 2016, pp. 48–69, doi:10.1111/sjp.12160.

Huvila, Isto. “Affective Capitalism of Knowing and the Society of Search Engine.” *Aslib Journal of Information Management*, vol. 68, no. 5, Sept. 2016, pp. 566–88, doi:10.1108/AJIM-11-2015-0178.

Introna, Lucas D., and Helen Nissenbaum. “Shaping the Web: Why the Politics of Search Engines Matters.” *The Information Society*, vol. 16, no. 3, 2000, pp. 169–185.

Ippolita. *The Dark Side of Google*. Translated by Patrice J.H Riemens, Institute of Network Cultures, 2013.

“Is Homosexuality Wrong?” *Debate.org*, www.debate.org/opinions/is-homosexuality-wrong. Accessed 28 Nov. 2017.

Jameson, Fredric. *Postmodernism: Or, the Cultural Logic of Late Capitalism*. Verso Books, 1992.

Jansen, Bernard J., et al. “Determining the Informational, Navigational, and Transactional Intent of Web Queries.” *Information Processing & Management*, vol. 44, no. 3, May 2008, pp. 1251–66, doi:10.1016/j.ipm.2007.07.015.

---. “Patterns of Query Reformulation During Web Searching.” *Journal of the American Society for Information Science and Technology*, vol. 60, no. 7, July 2009, pp. 1358–71, doi:10.1002/asi.21071.

Jansen, Bernard J., and Paulo R. Molina. “The Effectiveness of Web Search Engines for Retrieving Relevant Ecommerce Links.” *Information Processing & Management*, vol. 42, no. 4, July 2006, pp. 1075–98, doi:10.1016/j.ipm.2005.09.003.

Jansen, Bernard J., and Amanda Spink. “How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs.” *Information Processing & Management*, vol. 42, no. 1, Jan. 2006, pp. 248–63, doi:10.1016/j.ipm.2004.10.007.

Jarrett, Kylie. “A Database of Intention?” *Society of the Query Reader: Reflections on Web Search*, edited by René König and Miriam Rasch, Institute of Network Cultures, 2014, pp. 16–29.

---. *Feminism, Labour and Digital Media: The Digital Housewife*. Routledge, 2015.

Jarvis, Jeff. “Google Knowledge Graph Has More than 70 Billion Facts about People, Places, Things. + Language, Image, Voice Translation - @sundarpichai.” *Twitter*, @jefffarvis 4 Oct. 2016, twitter.com/jeffjarvis/status/783338071316135936.

---. *What Would Google Do?* HarperBusiness, 2011.

Jehovah’s Witnesses. “Is Homosexuality Wrong? | Young People Ask.” *JW.org Official Website of the Jehovah’s Witnesses*, www.jw.org/en/bible-teachings/teenagers/ask/is-homosexuality-wrong/. Accessed 29 Oct. 2017.

Jobin, Anna, and Olivier Glassey. “‘I Am Not a Web Search Result! I Am a Free Word.’ The Categorization and Commodification of ‘Switzerland’ by Google.” *Society of the Query Reader: Reflections on Web Search*, edited by René König and Miriam Rasch, Institute of Network Cultures, 2014, pp. 149–62.

Jones, Rhett. “RIP Google Instant Search, You Were Never Necessary.” *Gizmodo*, 26 July 2017, gizmodo.com/rip-google-instant-search-you-were-never-necessary-1797281069.

Kadouch, David. “Local Flavor for Google Suggest.” *Official Google Blog*, 31 Mar. 2009, googleblog.blogspot.com/2009/03/local-flavor-for-google-suggest.html.

Kao, Evelyn. “Making Search Results More Local and Relevant.” *Official Google Blog*, 27 Oct. 2017, www.blog.google/products/search/making-search-results-more-local-and-relevant/.

Kaplan, Frederic. “Linguistic Capitalism and Algorithmic Mediation.” *Representations*, vol. 127, no. 1, Aug. 2014, pp. 57–63, doi:10.1525/rep.2014.127.1.57.

Karapapa, Stavroula, and Maurizio Borghi. “Search Engine Liability for Autocomplete Suggestions: Personality, Privacy and the Power of the Algorithm.” *International Journal of Law and Information Technology*, vol. 23, no. 3, Sept. 2015, pp. 261–89, doi:10.1093/ijlit/eav009.

Kavoussi, Bonnie. “Google Chairman Eric Schmidt Defends Tax Dodge: ‘It’s Called Capitalism’.” *Huffington Post*, 13 Dec. 2012, www.huffingtonpost.com/2012/12/13/google-tax-dodge_n_2292077.html.

Keen, Andrew. *The Internet Is Not the Answer*. Atlantic Books, 2015.

Kelly, Kevin. *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*. Penguin Books, 2016.

Kitchin, Rob. “Thinking Critically about and Researching Algorithms.” *Information, Communication & Society*, vol. 20, no. 1, Jan. 2017, pp. 14–29, doi:10.1080/1369118X.2016.1154087.

Kittler, Friedrich. *Gramophone, Film, Typewriter*. Translated by Geoffrey Winthrop-Young and Michael Wutz, Stanford UP, 1999.

---. “Towards an Ontology of Media.” *Theory, Culture & Society*, vol. 26, no. 2–3, Mar. 2009, pp. 23–31, doi:10.1177/0263276409103106.

Kleinberg, Jon M. “Authoritative Sources in a Hyperlinked Environment.” *Journal of ACM*, vol. 46, no. 5, Sept. 1999, pp. 604–632, doi:10.1145/324133.324140.

Kluger, Jeffrey. “No Ben Carson, Homosexuality Is Not a Choice.” *Time*, Mar. 2015, time.com/3733480/ben-carson-gay-choice-science/.

König, René, and Miriam Rasch, editors. *Society of the Query Reader: Reflections on Web Search*. Institute of Network Cultures, 2014.

Kornai, András. “Digital Language Death.” *PLOS ONE*, vol. 8, no. 10, Oct. 2013, doi:10.1371/journal.pone.0077056.

Krebs, Viola, and Vincent Climent-Ferrando. “Languages, Cyberspace, Migrations.” *Net. Lang: Towards a Multilingual Cyberspace*, edited by Laurent Vannini and Hervé Le Crosnier, C&F éditions, 2012, pp. 229–48, unesdoc.unesco.org/images/0021/002166/216692e.pdf.

Kumar, B. T. Sampath, and J. N. Prakash. “Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo.” *Singapore Journal of Library & Information Management*, vol. 38, no. 1, 2009, pp. 124–137.

Kumar, Rakesh, et al. “Search Engines Evaluation.” *DESIDOC Journal of Library & Information Technology*, vol. 25, no. 2, 2005, doi:10.14429/dbit.25.2.3649.

Kurzweil, Raymond. *The Singularity Is Near*. Penguin Books, 2006.

LaFrance, Adrienne. “When Robots Hallucinate.” *The Atlantic*, Sept. 2015, www.theatlantic.com/technology/archive/2015/09/robots-hallucinate-dream/403498/.

Lamont, Ann. “What’s Wrong with Being Gay? Homosexual Behavior versus the Bible.” *Christian Answers*, 2004, christiananswers.net/q-aig/aig-c040.html.

Lamping, John, and Steven Baker. *Determining Query Term Synonyms within Query Context*. US7636714 B1, 22 Dec. 2009, www.google.com/patents/US7636714. Patent.

Landauer, Thomas K. “How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-Term Memory.” *Cognitive Science*, vol. 10, no. 4, 1986, pp. 477–493.

Langville, Amy N., and Carl D. Meyer. *Google’s PageRank and Beyond: The Science of Search Engine Rankings*. Princeton UP, 2012.

Lash, Scott. *Critique of Information*. Sage Publications, 2002.

Latour, Bruno. “Technology Is Society Made Durable.” *The Sociological Review*, vol. 38, no. S1, May 1990, pp. 103–31, doi:10.1111/j.1467-954X.1990.tb03350.x.

Lawrence, Steve, and C. Lee Giles. “Accessibility of Information on the Web.” *Nature*, vol. 400, no. 1, July 1999, pp. 107–09, doi:10.1038/21987.

Lazzarato, Maurizio. *Signs and Machines: Capitalism and the Production of Subjectivity*. MIT Press, 2014.

Lemon. *Idiots.win*. www.idiots.win. Accessed 1 Dec. 2017.

Leonhard, Gerd. *Technology vs. Humanity: The Coming Clash between Man and Machine*. Fast Future Publishing, 2016.

Levy, Steven. “How Google Is Remaking Itself as a ‘Machine Learning First.’” *Wired*, 22 June 2016, www.wired.com/2016/06/how-google-is-remaking-itself-as-a-machine-learning-first-company/.

---. *In the Plex: How Google Thinks, Works, and Shapes Our Lives*. Simon & Schuster, 2011.

Lewandowski, Dirk. “A Framework for Evaluating the Retrieval Effectiveness of Search Engines.” *Next Generation Search Engines: Advanced Models for Information Retrieval*, 2012, pp. 456–479.

---. “New Perspectives on Web Search Engine Research.” *Web Search Engine Research*, vol. 4, Emerald Group Publishing, 2012, pp. 1–16, doi:10.1108/S1876-0562(2012)002012a003.

---. “Why We Need an Independent Index of the Web.” *Society of the Query Reader: Reflections on Web Search*, edited by René König and Miriam Rasch, Institute of Network Cultures, 2014, pp. 50–58.

Lewandowski, Dirk, and Nadine Höchstötter. “Web Searching: A Quality Measurement Perspective.” *Web Search: Multidisciplinary Perspectives*, Springer, 2008, pp. 309–42, doi:10.1007/978-3-540-75829-7_14.

Lin, Yuri, et al. “Syntactic Annotations for the Google Books Ngram Corpus.” *Proceedings of the ACL 2012 System Demonstrations*, Association for Computational Linguistics, 2012, pp. 169–174, aclweb.org/anthology/P/P12/P12-3029.pdf.

Liu, Jennifer. “At a Loss for Words?” *Official Google Blog*, 25 Aug. 2008, googleblog.blogspot.com/2008/08/at-loss-for-words.html.

Llull, Ramon. *Doctor Illuminatus: A Ramon Llull Reader*. Edited and translated by Anthony Bonner, Princeton UP, 1994.

Long, Christopher. *Aristotle on the Nature of Truth*. Cambridge UP, 2010.

Lord, Albert Bates. *The Singer of Tales*. Harvard UP, 1960.

Love, Julia, and Kristina Cooke. “Google, Facebook Move to Restrict Ads on Fake News Sites.” *Reuters*, 15 Nov. 2016, uk.reuters.com/article/us-alphabet-advertising-idUKKBN1392MM.

Luka, Mary, and Mél Hogan. “Polluted and Predictive, in 133 Words.” *Society of the Query Reader: Reflections on Web Search*, edited by René König and Miriam Rasch, Institute of Network Cultures, 2014, pp. 239–55.

Lynch, John Patrick. *Aristotle’s School: A Study of a Greek Educational Institution*. University of California Press, 1972.

MacCormick, John. *Nine Algorithms That Changed the Future: The Ingenious Ideas That Drive Today’s Computers*. Princeton UP, 2012.

Machill, Marcel, et al. “Search-Engine Research: A European-American Overview and Systematization of an Interdisciplinary and International Research Field.” *Media, Culture & Society*, vol. 30, no. 5, Sept. 2008, pp. 591–608, doi:10.1177/0163443708094010.

Machill, Marcel, and Markus Beiler, editors. *Die Macht Der Suchmaschinen / The Power of Search Engines*. Herbert Von Halem Verlag, 2007.

Mahnke, Martina, and Emma Uprichard. “Algorithming the Algorithm.” *Society of the Query Reader: Reflections on Web Search*, edited by René König and Miriam Rasch, Institute of Network Cultures, 2014, pp. 256–70.

Marvin, Ginny. “Updated: A Visual History of Google Ad Labeling in Search Results.” *Search Engine Land*, 3 Mar. 2017, searchengineland.com/search-ad-labeling-history-google-bing-254332.

Mason, Paul. *PostCapitalism: A Guide to Our Future*. Penguin Books, 2016.

Mass Resistance. *What’s Wrong with Being “gay”? Here’s What They Don’t Tell You -- and It’s Really Disturbing*. 4 Feb. 2016, massresistance.org/docs/gen3/16a/its-killing-them/index.html.

Massumi, Brian. *Parables for the Virtual: Movement, Affect, Sensation*. Duke UP, 2002.

Mayer, Marissa. “Google Instant Launch Event.” *YouTube*, uploaded by Google, 8 Sept. 2010, www.youtube.com/watch?v=i0eMHRxlJ2c.

McMahon, Aine. “Brexit Prompts Increase in Queries about Irish Passports.” *The Irish Times*, 2016, www.irishtimes.com/news/social-affairs/brexit-prompts-increase-in-queries-about-irish-passports-1.2698203.

Metz, Cade. “AI Is Transforming Google Search. The Rest of the Web Is Next.” *Wired*, 4 Feb. 2016, www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, et al. “Distributed Representations of Words and Phrases and Their Compositionality.” *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119, papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.

Mikolov, Tomas, Ilya Sutskever, and Quoc Le. “Learning the Meaning Behind Words.” *Google Open Source Blog*, 14 Aug. 2013, opensource.googleblog.com/2013/08/learning-meaning-behind-words.html.

Miller, Claire Cain, and Nick Bilton. “How Pay-Per-Gaze Advertising Could Work with Google Glass.” *Bits Blog*, 20 Aug. 2013, bits.blogs.nytimes.com/2013/08/20/google-patents-real-world-pay-per-gaze-advertising/.

Mitchell, Amy, and Katerina Matsa. “The Declining Value of U.S. Newspapers.” *Pew Research Center*, 22 May 2015, www.pewresearch.org/fact-tank/2015/05/22/the-declining-value-of-u-s-newspapers/.

Mizzaro, Stefano. “Relevance: The Whole History.” *Journal of the American Society for Information Science*, vol. 48, no. 9, Sept. 1997, pp. 810–32, doi:10.1002/(SICI)1097-4571(199709)48:9<810::AID-ASI6>3.0.CO;2-U.

“Mobile Search Engine Market Share in the Russian Federation, July 2017.” *StatCounter Global Stats*, gs.statcounter.com/search-engine-market-share/mobile/russian-federation. Accessed 5 Oct. 2017.

“Mobile Search Engine Market Share in the United States of America, July 2017.” *StatCounter Global Stats*, gs.statcounter.com/search-engine-market-share/mobile/united-states-of-america. Accessed 31 Aug. 2017.

Moons, Michelle. “Palin on Paid Anti-Trump Protesters: ‘Not Even President Yet and Our Guy’s Already Creating Jobs.’” *Breitbart*, 1 July 2016, www.breitbart.com/2016-presidential-race/2016/07/01/palin-paid-trump-protesters-not-even-president-yet-guys-already-creating-jobs/.

Mowshowitz, Abbe, and Akira Kawaguchi. “Measuring Search Engine Bias.” *Information Processing and Management*, vol. 41, no. 5, Sept. 2005, pp. 1193–1205, doi:10.1016/j.ipm.2004.05.005.

Moz. *Google Algorithm Change History*. 8 Mar. 2017, moz.com/google-algorithm-change.

Nagle, Angela. *Kill All Normies: Online Culture Wars from 4chan and Tumblr to Trump and the Alt-Right*. Zero Books, 2017.

National Telecommunications and Information Administration. “A Nation Online: How Americans Are Expanding Their Use of the Internet.” *Washington, DC: US Department of Commerce*, 2002, www.digitale-chancen.de/transfer/downloads/MD222.pdf.

Neiwert, David. *Alt America: The Rise of the Radical Right in the Age of Trump*. Verso Books, 2017.

Nelson, Michael L., Joan A. Smith, et al. “Efficient, Automatic Web Resource Harvesting.” *Proceedings of the 8th Annual ACM International Workshop on Web Information and Data Management*, Association for Computing Machinery, 2006, pp. 43–50, doi:10.1145/1183550.1183560.

Nelson, Michael L., Herbert Van de Sompel, et al. “Mod_oai: An Apache Module for Metadata Harvesting.” *Cornell University Library arXiv*, Mar. 2005, arxiv.org/abs/cs/0503069.

Newport, Cal. *Deep Work: Rules for Focused Success in a Distracted World*. Piatkus, 2016.

Ng, Andrew. “A Massive Google Network Learns to Identify — Cats.” *All Things Considered,* interview by Audie Cornish, National Public Radio, 26 June 2012, www.npr.org/2012/06/26/155792609/a-massive-google-network-learns-to-identify.

Nielsen, Jakob. “F-Shaped Pattern For Reading Web Content.” *Nielsen Norman Group*, 17 Apr. 2006, www.nngroup.com/articles/f-shaped-pattern-reading-web-content/.

Norvig, Peter. “How to Write a Spelling Corrector.” *Norvig.com*, Aug. 2016, www.norvig.com/spell-correct.html.

O'Hare, Ryan. "Google Search Spike Suggests Many People Don't Know Why They Voted for Brexit." *Mail Online*, 24 June 2016, www.dailymail.co.uk/sciencetech/article-3657997/Britain-s-Google-searches-hint-people-didn-t-know-voting-for.html.

Olanoff, Drew. "Zuckerberg On Building a Search Engine: Facebook Is Pretty Uniquely Positioned, At Some Point We'll Do It." *TechCrunch*, 11 Sept. 2012, social.techcrunch.com/2012/09/11/zuckerberg-we-have-a-team-working-on-search/.

Ong, Walter J. *Orality and Literacy: The Technologizing of the Word*. Routledge, 1982.

Oppenheim, Charles, et al. "The Evaluation of WWW Search Engines." *Journal of Documentation*, vol. 56, no. 2, Apr. 2000, pp. 190–211, doi:10.1108/00220410010803810.

Özmutlu, H. Cenk, et al. "Neural Network Applications for Automatic New Topic Identification on Excite Web Search Engine Data Logs." *Proceedings of the Association for Information Science and Technology*, vol. 41, no. 1, 2004, pp. 310–316, doi:10.1002/meet.1450410137.

Page, Danny. "Stop Using Google Trends." *Medium*, 24 June 2016, medium.com/@dannypage/stop-using-google-trends-a5014dd32588#.vop9ejer4.

Page, Lawrence, et al. "The PageRank Citation Ranking: Bringing Order to the Web." *Stanford InfoLab*, 1999, ilpubs.stanford.edu:8090/422.

Palacios, Rodrigo. "Autocomplete - an Adult and Kid Friendly Exercise in Creating a Predictive Program." *GitHub*, 14 May 2015, github.com/rodricios/autocomplete.

Papoutsaki, Alexandra, et al. "SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search." *CHIIR '17 Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 2017, pp. 17–26, doi:10.1145/3020165.3020170.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books, 2012.

Parkinson, Hannah. “Click and Elect: How Fake News Helped Donald Trump Win a Real Election.” *The Guardian*, 14 Nov. 2016, www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies.

Parry, Milman. *L’épithète Traditionnelle Dans Homère: Essai Sur Un Problème De Style Homérique*. Les belles lettres, 1928.

Pasquale, Frank. “From Holocaust Denial to Hitler Admiration, Google’s Algorithm Is Dangerous.” *Huffington Post*, 6 Feb. 2017, www.huffingtonpost.com/entry/holocaust-google-algorithm_us_587e8628e4b0c147f0bb9893.

Pasquinelli, Matteo. “Google’s PageRank Algorithm: Diagram of Cognitive Capitalism and the Rentier of the Common Intellect.” *Deep Search. The Politics of Search Beyond Google*, edited by Konrad Becker and Felix Stalder, Studien Verlag, 2009, pp. 152–62.

Pass, Greg, et al. “A Picture of Search.” *Proceedings of the 1st International Conference on Scalable Information Systems*, Association for Computing Machinery, 2006, doi:10.1145/1146847.1146848.

Peifer, Karl-Nikolaus. “Google’s Autocomplete Function – Is Google a Publisher or Mere Technical Distributor? German Federal Supreme Court, Judgment of May 14, 2012 – Case No. Vi Zr 269/12 – Google Autocomplete.” *Queen Mary Journal of Intellectual Property*, vol. 3, no. 4, Oct. 2013, pp. 318–24, doi:10.4337/qmjip.2013.04.05.

Peters, John Durham. *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. University of Chicago Press, 2015.

Pickering, Andrew. *The Mangle of Practice: Time, Agency, and Science*. University of Chicago Press, 1995.

Pickering, Andrew, and Adam Stephanides. “Constructing Quaternions: On the Analysis of Conceptual Practice.” *Science as Practice and Culture*, edited by Andrew Pickering, University of Chicago Press, 1992, pp. 139–67.

Pitkow, James, et al. “Personalized Search.” *Communications of the Association for Computing Machinery*, vol. 45, no. 9, Sept. 2002, pp. 50–55.

Plato. “Meno.” *Plato: Complete Works*, edited by John M. Cooper, Hackett Publishing Company, 1997, pp. 870–97.

---. “Phaedrus.” *Plato: Complete Works*, edited by John M. Cooper, Hackett Publishing Company, 1997, pp. 506–56.

@pollyp1. “Top 5 Search Suggestions for Female Scientists Often Include ‘husband’. WTF?” *Twitter*, 20 May 2016, twitter.com/pollyp1/status/733636162263351296?lang=en-gb.

Popyer, Kacy. “Cache-22: The Fine Line Between Information and Defamation in Google’s Autocomplete Function.” *Cardozo Arts and Entertainment Law Journal*, vol. 34, 2016, pp. 835–61.

Prado, Daniel. “Language Presence in the Real World and Cyberspace.” *Net. Lang: Towards a Multilingual Cyberspace*, edited by Laurent Vannini and Hervé Le Crosnier, C&F éditions, 2012, pp. 38–51, unesdoc.unesco.org/images/0021/002166/216692e.pdf.

Purcell, Kristin, et al. “Search Engine Use 2012.” *Pew Internet & American Life Project*, Mar. 2012, www.pewinternet.org/2012/03/09/search-engine-use-2012/.

Racer Nation Information. “MBA Students Use Google, Stukent Digital Marketing Tools.” *Racer Nation Information*, 20 May 2016,

www.racernationinformation.com/2016/05/19/mba-students-use-google-stukent-digital-marketing-tools/.

Raff, Adam. “Search, but You May Not Find.” *The New York Times*, 27 Dec. 2009, www.nytimes.com/2009/12/28/opinion/28raff.html.

Rather, Rafiq A., and S. M. Shafi. “Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology.” *Webology*, vol. 2, no. 2, Aug. 2005, www.webology.org/2005/v2n2/a12.html.

Reilly, Paul. “‘Googling’ Terrorists: Are Northern Irish Terrorists Visible on Internet Search Engines?” *Web Search: Multidisciplinary Perspectives*, Springer, 2008, pp. 151–75, doi:10.1007/978-3-540-75829-7_10.

Roberts, Jeff. “Brits Scramble to Google ‘What Is the EU?’ Hours After Voting to Leave It.” *Fortune*, 24 June 2016, fortune.com/2016/06/24/brexit-google-trends/.

Rogers, Simon. “What Is Google Trends Data — and What Does It Mean?” *Medium*, 1 July 2016, medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8#.n9jt853qg.

Rose, Daniel E. “Reconciling Information-Seeking Behavior with Search User Interfaces for the Web.” *Journal of the American Society for Information Science and Technology*, vol. 57, no. 6, Apr. 2006, pp. 797–99, doi:10.1002/asi.20295.

Rose, Daniel E., and Danny Levinson. “Understanding User Goals in Web Search.” *Proceedings of the 13th International Conference on World Wide Web*, Association for Computing Machinery, 2004, pp. 13–19, doi:10.1145/988672.988675.

Rose, Steven. “Memories Are Made of This.” *Memory: Histories, Theories, Debates*, edited by Susannah Radstone and Bill Schwarz, Fordham UP, 2010, pp. 198–208.

Ruban, S., and S. Behin Sam. “An Algorithmic Query Refinement Model Based on Query Classification.” *Indian Journal of Science and Technology*, vol. 10, no. 18, June 2017, www.indjst.org/index.php/indjst/article/view/114051.

SadBotTrue. “Chapter 39. The Fake Media.” *SadBotTrue*, sadbottrue.com/article/72/. Accessed 9 June 2017.

Saffer, Dan. *Microinteractions: Designing with Details*. O’Reilly Media, 2013.

@sammich_BLT. “The Lack of Context Really Is Astounding.” *Twitter*, 24 June 2016, twitter.com/Sammich_BLT/status/746365855441657856.

Sanz, Esteve, and Juraj Stančík. “Your Search – ‘Ontological Security’ – Matched 111,000 Documents: An Empirical Substantiation of the Cultural Dimension of Online Search.” *New Media & Society*, vol. 16, no. 2, Mar. 2014, pp. 252–70, doi:10.1177/1461444813481198.

Scanlan, John. *Memory: Encounters with the Strange*. Reaktion Books, 2013.

Schamber, Linda, et al. “A Re-Examination of Relevance.” *Information Processing and Management*, vol. 26, no. 6, 1990, pp. 755–76, doi:10.1016/0306-4573(90)90050-C.

Schmidt, Eric, and Jared Cohen. *The New Digital Age: Reshaping the Future of People, Nations and Business*. John Murray, 2013.

Schmidt, Eric, and Jonathan Rosenberg. *How Google Works*. Grand Central Publishing, 2017.

Schuler, Douglas. “Reports of the Close Relationship Between Democracy and the Internet.” *Democracy and Digital Media*, May 1998, web.mit.edu/m-i-t/articles/schuler.html.

Schwartz, Barry. “Google Drops Another Search Operator: Tilde for Synonyms.” *Search Engine Land*, 24 June 2013, searchengineland.com/google-drops-another-search-operator-tilde-for-synonyms-164403.

---. “Google Has Dropped Google Instant Search.” *Search Engine Land*, 26 July 2017, searchengineland.com/google-dropped-google-instant-search-279674.

Schwartz, Bill. “Memory, Temporality, Modernity.” *Memory: Histories, Theories, Debates*, edited by Susannah Radstone and Bill Schwarz, Fordham UP, 2010, pp. 41–60.

“Search Engine Market Share Worldwide, July 2017.” *StatCounter Global Stats*, gs.statcounter.com/search-engine-market-share. Accessed 5 Oct. 2017.

Selyukh, Alina. “After Brexit Vote, Britain Googles ‘What Is The EU?’” *National Public Radio*, 24 June 2016, www.npr.org/sections/alltechconsidered/2016/06/24/480949383/britains-google-searches-for-what-is-the-eu-spike-after-brexit-vote.

Shestakov, Denis, and Tapio Salakoski. “Sampling the National Deep Web.” *Database and Expert Systems Applications*, Springer, 2011, pp. 331–40, doi:10.1007/978-3-642-23088-2_24.

Shi, Zhiwei, et al. “Using Global Statistics to Rank Retrieval Systems Without Relevance Judgments.” *Intelligent Information Processing V*, Springer, 2010, pp. 183–92, doi:10.1007/978-3-642-16327-2_24.

Shivakumar, Shiva. “New ‘Google Sitemaps’ Web Page Feed Program.” *Search Engine Watch* Interview by Danny Sullivan, 8 June 2005, web.archive.org/web/20050608000513/blog.searchenginewatch.com:80/blog/050602-195224.

Silverman, Craig. “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook.” *BuzzFeed*, 16 Nov. 2016, www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook.

Silverman, Craig, and Lawrence Alexander. “How Teens in The Balkans Are Duping Trump Supporters with Fake News.” *BuzzFeed*, 3 Nov. 2016, www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo.

Silvestri, Fabrizio. “Mining Query Logs: Turning Search Usage Data into Knowledge.” *Foundations and Trends in Information Retrieval*, vol. 4, no. 1–2, 2010, pp. 1–174, doi:10.1561/1500000013.

Simonite, Tom. “Google’s Loon Balloons Are Ready to Deliver Cheap Internet.” *MIT Technology Review*, 2015, www.technologyreview.com/s/534986/project-loon/.

Singel, Ryan. “Google Debuts ‘Instant’ Search (Updated).” *Wired*, www.wired.com/2010/09/google-search-update-2/. Accessed 15 Aug. 2017.

Singhal, Amit. “Introducing the Knowledge Graph: Things, Not Strings.” *Official Google Blog*, 16 May 2012, googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.

---. “Microsoft’s Bing Uses Google Search Results—and Denies It.” *Official Google Blog*, googleblog.blogspot.com/2011/02/microsofts-bing-uses-google-search.html. Accessed 26 Aug. 2017.

Solomon, Brian. “Google Just Passed Apple as The World’s Most Valuable Company.” *Forbes*, 1 Feb. 2016, www.forbes.com/sites/briansolomon/2016/02/01/google-just-passed-apple-as-the-worlds-most-valuable-company/.

---. “Google Passed Apple as The World’s Most Valuable Company (Again).” *Forbes*, 12 May 2016, www.forbes.com/sites/briansolomon/2016/05/12/google-passed-apple-as-the-worlds-most-valuable-company-again/.

Solove, Daniel J. *The Digital Person: Technology and Privacy in the Information Age*. NYU Press, 2006.

Sparrow, et al. “Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips.” *Science*, vol. 333, 2011, pp. 776–78.

Sparrow, Betsy, and Ljubica Chatman. “Social Cognition in the Internet Age: Same As It Ever Was?” *Psychological Inquiry*, vol. 24, no. 4, Oct. 2013, pp. 273–92, doi:10.1080/1047840X.2013.827079.

Spink, A., et al. “From E-Sex to E-Commerce: Web Search Changes.” *Computer*, vol. 35, no. 3, Mar. 2002, pp. 107–09, doi:10.1109/2.989940.

Spink, Amanda, and Michael Zimmer, editors. *Web Search: Multidisciplinary Perspectives*. Vol. 14, Springer, 2008, doi:10.1007/978-3-540-75829-7.

Stalder, Felix, and Christine Mayer. “The Second Index. Search Engines, Personalization and Surveillance.” *Deep Search: The Politics of Search Beyond Google*, edited by Konrad Becker and Felix Stalder, Studien Verlag, 2009.

Statt, Nick. “Google Will Stop Showing Search Results as You Type Because It Makes No Sense on Mobile.” *The Verge*, 26 July 2017, www.theverge.com/2017/7/26/16034844/google-kills-off-instant-search-for-mobile-consistency.

Sterling, Greg. “Report: Nearly 60 Percent of Searches Now from Mobile Devices.” *Search Engine Land*, 3 Aug. 2016, searchengineland.com/report-nearly-60-percent-searches-now-mobile-devices-255025.

Stopera, Matt, and Lauren Yapalater. “Why Being Gay Is Better Than Being Straight.” *BuzzFeed*, 14 June 2016, www.buzzfeed.com/mjs538/why-being-gay-is-better-than-being-straight.

Stross, Randall. *Planet Google: How One Company Is Transforming Our Lives*. Atlantic Books, 2009.

Subramanian, Samanth. “Inside the Macedonian Fake-News Complex.” *Wired*, 15 Feb. 2017, www.wired.com/2017/02/veles-macedonia-fake-news/.

Suen, Brennan, et al. “Websites Peddling Fake News Still Using Google Ads Nearly a Month After Google Announced Ban.” *Media Matters for America*, 14 Dec. 2016, www.mediamatters.org/research/2016/12/14/websites-peddling-fake-news-still-using-google-ads-nearly-month-after-google-announced-ban/214811.

Sullivan, Danny. “Dear Bing, We Have 10,000 Ranking Signals to Your 1,000. Love, Google.” *Search Engine Land*, 11 Nov. 2010, searchengineland.com/bing-10000-ranking-signals-google-55473.

---. “FAQ: All About the Google RankBrain Algorithm.” *Search Engine Land*, 23 June 2016, searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440.

---. “Google: Bing Is Cheating, Copying Our Search Results.” *Search Engine Land*, 1 Feb. 2011, searchengineland.com/google-bing-is-cheating-copying-our-search-results-62914.

---. “Google Now Handles at Least 2 Trillion Searches Per Year.” *Search Engine Land*, 24 May 2016, searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247.

---. “Google Uses RankBrain for Every Search, Impacts Rankings of ‘Lots’ of Them.” *Search Engine Land*, 23 June 2016, searchengineland.com/google-loves-rankbrain-uses-for-every-search-252526.

---. “Schmidt: Listing Google’s 200 Ranking Factors Would Reveal Business Secrets.” *Search Engine Land*, 17 Sept. 2010, searchengineland.com/schmidt-listing-googles-200-ranking-factors-would-reveal-business-secrets-51065.

Sun, Yang. *A Comprehensive Study of the Regulation and Behavior of Web Crawlers*. Dissertation, Pennsylvania State University 2008, etda.libraries.psu.edu/catalog/9230.

superhiffy. “Comment on ‘Crystal Fairy – Official Trailer.’” *YouTube*, 3 Jan. 2014, www.youtube.com/watch?v=E7nInvUzAI8.

Sutton, John. *Philosophy and Memory Traces: Descartes to Connectionism*. Cambridge UP, 2008.

Sutton, John, et al. “Memory and Cognition.” *Memory: Histories, Theories, Debates*, edited by Susannah Radstone and Bill Schwarz, Fordham UP, 2010, pp. 209–26.

Talon, Emmanuelle. “L’arabe, Une « Langue de France » Sacrifiée.” *Le Monde Diplomatique*, 1 Oct. 2012, www.monde-diplomatique.fr/2012/10/TALON/48275.

Teevan, Jaime, et al. “Personalizing Search via Automated Analysis of Interests and Activities.” *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, 2005, pp. 449–456, doi:10.1145/1076034.1076111.

TFP Student Action. “10 Reasons Why Homosexual ‘Marriage’ Is Harmful and Must Be Opposed.” *TFP Student Action*, www.tfpstudentaction.org/blog/10-reasons-why-homosexual-marriage-is-harmful-and-must-be-opposed. Accessed 29 Oct. 2017.

Thakur, Shashi. “Feed Your Need to Know.” *Official Google Blog*, 19 July 2017, www.blog.google:443/products/search/feed-your-need-know/.

Truong, Alice. “Could Google Glass Track Your Emotional Response to Ads?” *Fast Company*, 13 Aug. 2013, www.fastcompany.com/3015654/could-google-glass-track-your-emotional-response-to-ads.

@tuuliel. “Dear Fellow Scientists. The next Time You’re Going to Google for Info of the Husband or Wedding of a Female Colleague, Think Again. 1/6 ... I Want a Scientific

Community That Treats and Thinks of Women as Professionals, Not as Objects of Desire or Wives of Someone! 6/6." *Twitter*, 15 July 2017, twitter.com/tuuliel/status/886272533154103296?lang=en.

Tynan, Dan. "How Facebook Powers Money Machines for Obscure Political 'News' Sites." *The Guardian*, 24 Aug. 2016, www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump.

Ulmer, Gregory L. *Internet Invention: From Literacy to Electracy*. Pearson, 2002.

UN Women. "Un Women Ad Series Reveals Widespread Sexism." *UN Women*, 21 Oct. 2013, www.unwomen.org/en/news/stories/2013/10/women-should-ads.

Vaidhynathan, Siva. *The Googlization of Everything (And Why We Should Worry)*. 2nd revised edition, University of California Press, 2012.

Van Couvering, Elizabeth. "The History of the Internet Search Engine: Navigational Media and the Traffic Commodity." *Web Search: Multidisciplinary Perspectives*, edited by Amanda Spink and Michael Zimmer, vol. 14, Springer, 2008, pp. 177–206.

van Gurp, Marc. "Google's Autocomplete Feature Is Shocking Too About Black Men." *Osocio*, 17 Nov. 2013, osocio.org/message/googles-autocomplete-feature-is-shocking-too-about-black-men/.

---. "Romanians Are Smart." *Osocio*, 23 Nov. 2011, osocio.org/message/google-romanians-are-stupid/.

---. "The Shocking Answers from Google's Autocomplete Feature on Sexism." *Osocio*, 20 Oct. 2013, osocio.org/message/the-shocking-answers-from-googles-autocomplete-feature-on-sexism/.

Virno, Paolo. *A Grammar of the Multitude*. Semiotext[e], 2004.

Virno, Paolo, and Michael Hardt, editors. *Radical Thought in Italy: A Potential Politics*. University of Minnesota Press, 2006.

Walton, Mark. “Many UK Voters Didn’t Understand Brexit, Google Searches Suggest.” *Ars Technica*, 24 June 2016, arstechnica.com/tech-policy/2016/06/brexit-google-search-trends-tech/.

Wegner, Daniel, et al. “Transactive Memory in Close Relationships.” *Journal of Personality and Social Psychology*, vol. 61, no. 6, Dec. 1991, pp. 923–29.

Wegner, Daniel M., et al. “Cognitive Interdependence in Close Relationships.” *Compatible and Incompatible Relationships*, Springer, 1985, pp. 253–276.

White, Jeremy. “The Violence of Charlottesville Has Been a Clarifying Moment for Silicon Valley.” *The Independent*, 18 Aug. 2017, www.independent.co.uk/news/world/americas/tech-charlottesville-daily-stormer-cloudflare-google-paypal-apple-godaddy-a7901491.html.

Wiggins, Martin. “Shakespeare Lives - Who Wrote Shakespeare?” *BBC*, www.bbc.co.uk/programmes/articles/211LBPTmBYp2rbh4bSQlSTS/who-wrote-shakespeare. Accessed 20 Nov. 2017.

“Wikipedia:About.” *Wikipedia*, en.wikipedia.org/wiki/Wikipedia:About.

Williams, Raymond. “Advertising: The Magic System.” *Problems in Materialism and Culture*, Verso Books, 1980, pp. 170–95.

Willis, Amy. “Britons Are Googling ‘what is the EU’ After Voting to Leave It.” *The Metro*, 24 June 2016, metro.co.uk/2016/06/24/britons-are-googling-what-is-the-eu-after-voting-to-leave-it-5965238/.

Wilson, Theodore. “Google Searches for ‘What is the EU?’ and ‘How to Emigrate’ Spike after Devastating Brexit Vote.” *The Mirror*, 24 June 2016, www.mirror.co.uk/tech/google-searches-what-eu-spike-8274842.

Wingfield, Nick, et al. “Google and Facebook Take Aim at Fake News Sites.” *The New York Times*, 14 Nov. 2016, www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html.

Winkler, Mathis. “Germany Pulls Away from Quaero Search-Engine Project.” *Deutsche Welle*, 21 Dec. 2006, www.dw.com/en/germany-pulls-away-from-quaero-search-engine-project/a-2287489.

Winthrop-Young, Geoffrey. *Kittler and the Media*. Polity, 2010.

Wired. “WIRED’s Autocomplete Interviews.” *YouTube*, uploaded by WIRED www.youtube.com/playlist?list=PLibNZv5Zd0dwjZFCTVZ8QdKq194CkwXjo. Accessed 28 Oct. 2017.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. 2nd edition, Routledge, 2001.

Wohlsen, Marcus. “The PC’s Death Might Also Mean the Web’s Demise.” *Wired*, 13 Jan. 2014, www.wired.com/2014/01/death-pc-also-mean-end-web/.

Wright, Alex. *Glut: Mastering Information through the Ages*. Cornell UP, 2008.

Wright, Joshua D. “Defining and Measuring Search Bias: Some Preliminary Evidence.” *International Center for Law and Economics*, George Mason University Law and Economics Research Paper Series, Nov. 2011, papers.ssrn.com/sol3/papers.cfm?abstract_id=2004649.

Yates, Frances. *Giordano Bruno and the Hermetic Tradition*. Routledge and Kegan Paul, 1964.

---. *The Art of Memory*. Routledge and Kegan Paul, 1966.

Zhang, Jin, and Wei Fei. “Search Engines’ Responses to Several Search Feature Selections.” *International Information & Library Review*, vol. 42, no. 3, Sept. 2010, pp. 212–25, doi:10.1080/10572317.2010.10762866.

Zhang, Ying, et al. “Time Series Analysis of a Web Search Engine Transaction Log.” *Information Processing & Management*, vol. 45, no. 2, Mar. 2009, pp. 230–45, doi:10.1016/j.ipm.2008.07.003.

Zimmer, Michael. “The Gaze of the Perfect Search Engine: Google as an Infrastructure of Dataveillance.” *Web Search: Multidisciplinary Perspectives*, Springer, 2008, pp. 77–99, doi:10.1007/978-3-540-75829-7_6.